JMIR Human Factors

Impact Factor (2023): 2.6
Volume 12 (2025) ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

Contents

Original Papers

Reviews

Original paper

Letter to the Editor

Research Letter

A Health Professional Mentorship Platform to Improve Equitable Access to Abortion: Development, Usability, and Content Evaluation

Abdul-Fatawu Abdulai[1], PhD; Cam Duong[1], PhD; Eleni Stroulia[2], PhD; Efrat Czerniak[1], PhD; Rachel Chiu[3], BSc; Aashay Mehta[1], MA; Ken Koike[4], MPH; Wendy V Norman[5], FCFP, DTM&H, MD, MHSc

[1]School of Nursing, University of British Columbia, Vancouver, BC, Canada

[2]Computing Science, Faculty of Science, University of Alberta, Edmonton, AB, Canada

[3]Cognitive Sciences, University of British Columbia, Vancouver, BC, Canada

[4]The Society of Obstetricians and Gynaecologists of Canada, Ottawa, ON, Canada

[5]Department of Family Practice, Faculty of Medicine, University of British Columbia, Vancouver, ON, Canada

Corresponding Author:
Abdul-Fatawu Abdulai, PhD
School of Nursing, University of British Columbia, Vancouver, BC, Canada

Abstract

Background: Access to safe abortion care is a reproductive right for all individuals across Canada. Underserved populations are overrepresented among those with unintended pregnancies and particularly those seeking abortion. Yet, few resources exist to help health care and allied helping professionals provide culturally competent and gender-affirming abortion care to such a population group.

Objective: This project aimed to redesign and adapt an existing subscription-based medication abortion mentorship platform into a culturally appropriate and gender-affirming open-access website of curated health professional resources to promote equitable, accessible, high-quality abortion care, particularly for underserved populations.

Methods: We drew on a user-centered design framework to redesign the web platform in 5 iterative phases. Health care and allied helping professionals were engaged in each stage of the development process including the initial design of the platform, curation of the resources, review of the content, and evaluation of the wireframes and the end product.

Results: This project resulted in an open-access bilingual (English and French) web-based platform containing comprehensive information and resources on abortion care for health care providers (physicians, nurse practitioners, and pharmacists) and allied helping professionals (midwives, medical officers, community workers, and social workers). The website incorporated information on clinical, logistical, and administrative guidance, including culturally competent and gender-affirming toolkits that could equip health care professionals with the requisite knowledge to provide abortion care for underserved populations.

Conclusions: This platform contains resources that can increase the competencies of health care professionals to initiate and sustain culturally and contextually appropriate abortion care for underserved groups while clarifying myths and misconceptions that often militate against initiating abortion. Our resource also has the potential to support equitable access to high-quality abortion care, particularly for those among underserved populations who may have the greatest unmet need for abortion services yet face the greatest barriers to accessing care.

(JMIR Hum Factors 2025;12:e63364) doi:10.2196/63364

KEYWORDS

medication abortion; mifepristone; web-based platform; user-centered design; underserved populations; abortion; equitable; accessibility; open-access website; gender-affirming; user-centered; Canada; unwanted pregnancy; framework

Introduction

Access to abortion is a fundamental human right and a critical component of sexual and reproductive health care [1]. Each year, approximately 56 million abortions occur across the globe—translating to 35 abortions per 1000 women. In Canada, nearly half (~40%) of pregnancies are unplanned, and one-third of women and pregnancy-capable people will have at least 1 abortion in their lifetime [2,3]. Medication abortion using mifepristone is the international gold standard for first-trimester medication abortions and currently accounts for over 50% of all abortions in most European countries and the United States. Canada's unique regulations allow nonphysician prescribers to prescribe and dispense the medication and allow people without

access to ultrasound to access medication abortion services. Despite the loosening of restrictions in Canada, timely access to safe abortion services can be difficult for historically, persistently, or systemically marginalized (HPSM) populations including racialized groups, migrants, Indigenous people, people with disabilities, homeless and underhoused people, sex workers, two-spirit, lesbian, gay, bisexual, transgender, queer, intersex, and gender-diverse (2SLGBTQI+) people, youths, and people living in rural and remote areas [4,5]. For instance, Indigenous people in Canada experience more abortion access barriers than non-Indigenous Canadian people [6]. In addition, approximately 18% of people in Canada traveled more than 100 km to access abortion services [7]. While most clinicians in Canada indicated providing abortion services to underserved populations, the majority of them did not have the appropriate cultural training to provide care to underserved populations. For instance, a recent study suggests that most clinicians in Canada who provide abortion services to underserved populations (91.2%) reported not receiving any form of training for providing care to such populations [8]. A lack of training or mentorship in culturally safe and gender-affirming care would further exacerbate barriers to accessing abortion services that already bedevil underserved populations including Indigenous people [9,10].

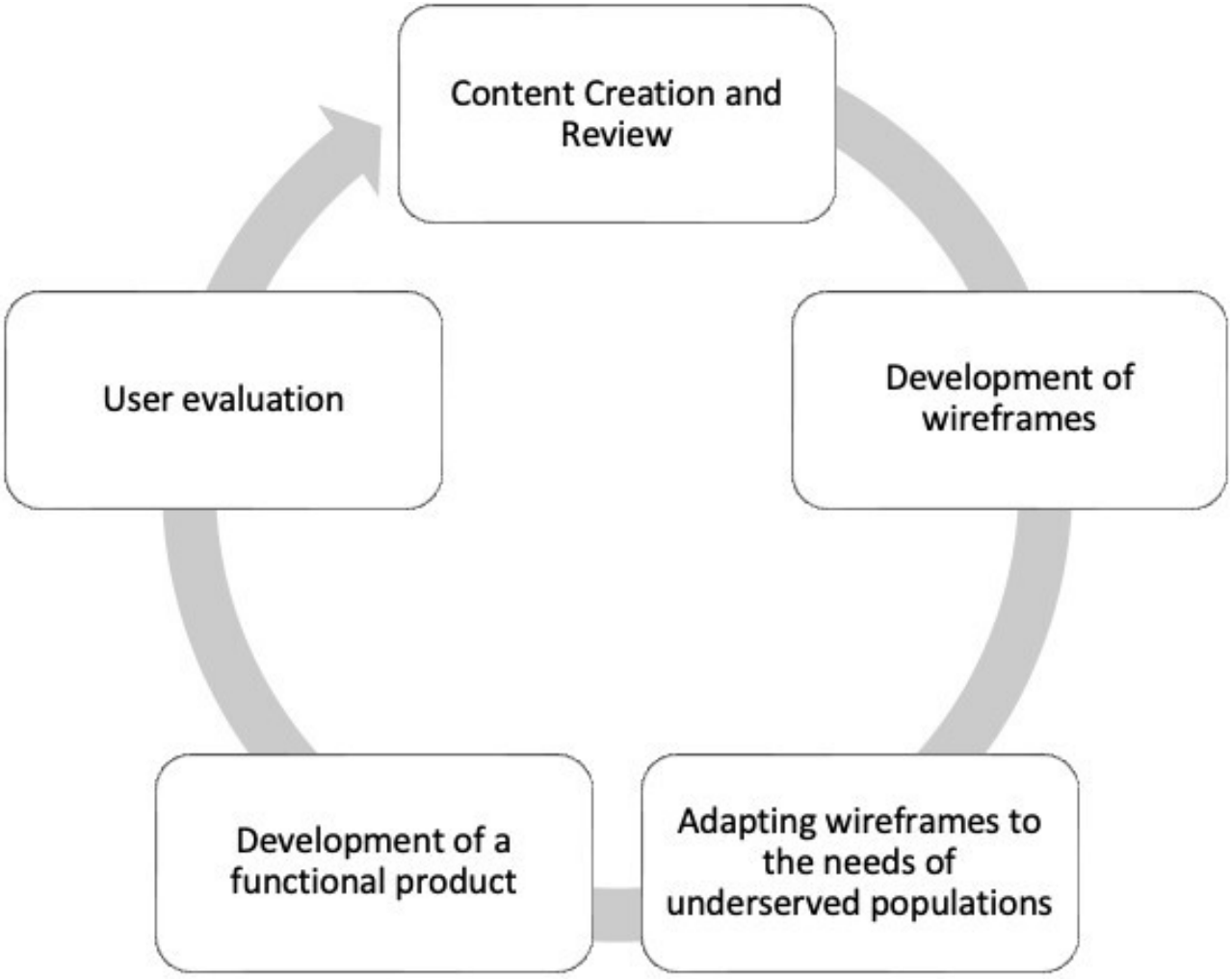
To provide culturally safe and gender-affirming abortion services, there is an urgent need to develop, evaluate, and implement training and mentorship resources that will equip health care professionals in providing equitable abortion services to underserved populations in Canada. In this paper, we reported on the processes we adopted in redesigning our already-established web-based community of practice platform into an open-access website. We also highlighted on the considerations for developing a gender-affirming, culturally appropriate, and inclusive web platform to support health professionals in providing abortion to equity-deserving populations. With the internet increasingly becoming a major source of information on abortion [11], an open-access web platform could not only help in disseminating evidence-based abortion resources to all health care professionals but also help in dispelling the myths and misconceptions that often fuel abortion-related stigmas. Furthermore, an open-access web platform could assist in countering abortion-related misinformation or disinformation that is often targeted at health professionals by antichoice groups.

Methods

Study Design

We used a mixed methods user-centered design approach in developing the content and structure of the website [12]. In compliance with the user-centered design process, we engaged with the relevant stakeholders in Canada (health professional associations and abortion advocacy organizations) in 5 iterative design phases. The five phases include (1) content creation and review, (2) development of wireframes, (3) adaptation of wireframes to end-user needs, (4) development of a functional product, and (5) user evaluation. We adopted a user-centered design process to emphasize the clinician-centeredness of the website while revealing the knowledge gaps that need to be addressed to create an inclusive web platform. Figure 1 shows a schematic presentation of the website's development process.

Figure 1. A schematic depiction of the website development process.



Ethical Considerations

This project was approved by the University of British Columbia Behavioral Research Ethics Board (approval H22-03342). All participants who took part in the interviews were provided with CAD $150 (US $103.34) honoraria.

Phase 1: Content Creation and Review

Overview

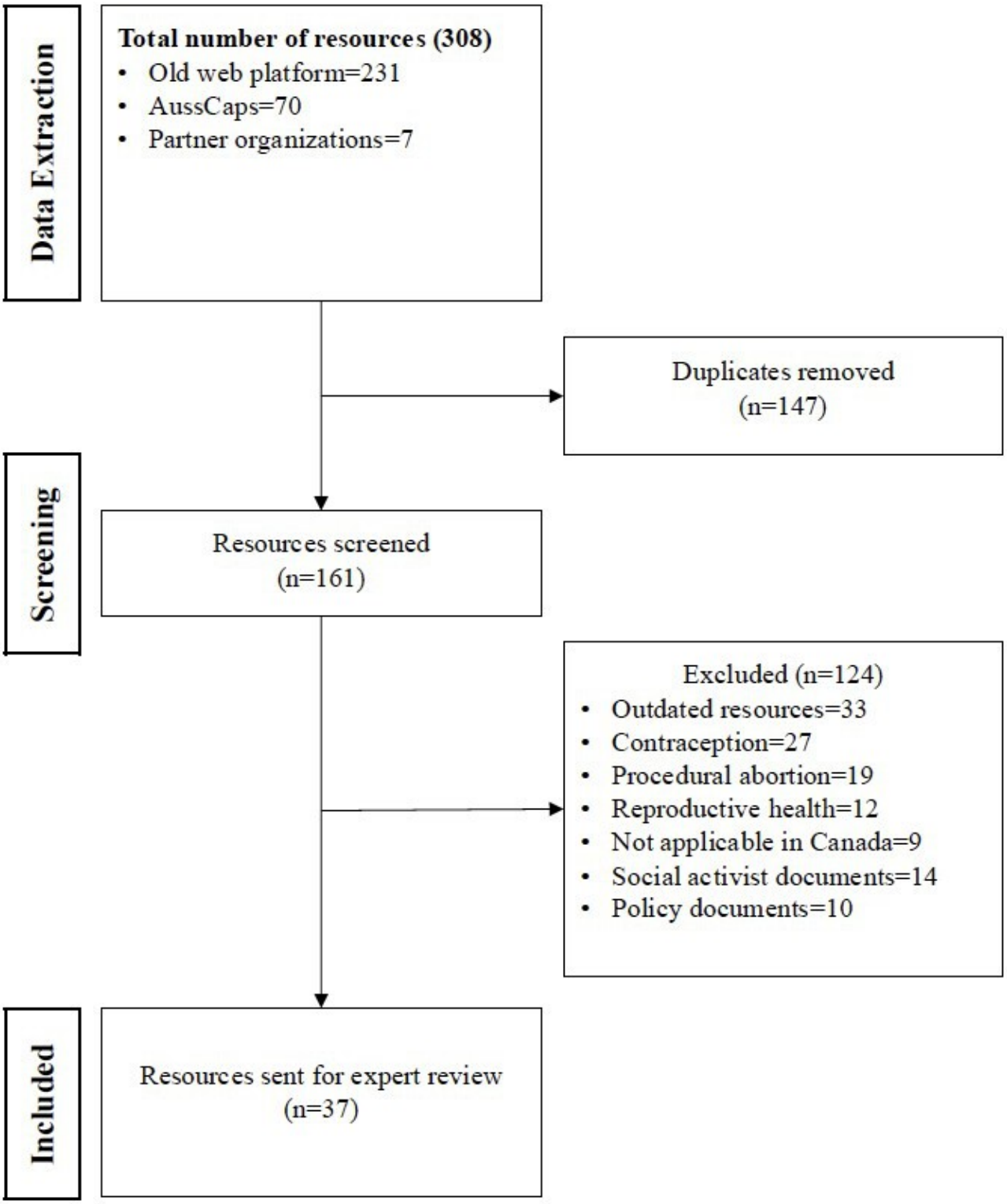
The content creation phase comprised 2 steps to ensure that the content is up-to-date and reliable. These include (1) the content curation process, in which we searched for resources that are relevant to medication abortion, and (2) the content review process.

Content Curation

We actively engaged with health professional associations, health regulatory agencies, and sexual and reproductive health advocacy organizations across Canada to curate relevant content for the website. To ensure high-quality, appropriate, evidence-based materials specific to Canadian health care professionals, we focused our content curation efforts on sources within Canada as well as trusted and reliable international evidence on medication abortion. The main sources for curating our content include the national health professional organizations (Society of Obstetricians and Gynaecologists of Canada, Canadian Pharmacy Association, Canadian Association of Midwives, and Canadian Association of Schools of Nursing) and advocacy groups (Action Canada for Sexual Health and Rights and National Abortion Federation) as well as our already-established web-based community of practice platform [13]. We examined relevant resources from a partner website in Australia (the Australian Contraception and Abortion Primary Care Practitioner Support Network) [14]. We limited our focus to medication abortion to enhance relevance and usability for clinicians looking to adopt, improve, and sustain the practice of this newly available service in Canada. After identifying the data sources and establishing a working relationship with stakeholders, we undertook focused data extraction from May 2023 to November 2023. The data extraction process yielded a total of 308 documents detailing resources on medication abortion. These were compiled into a Microsoft Excel sheet, and 147 duplicates were removed. Following this, we conducted an initial screening of the resources by examining the titles, abstracts, headings, and subheadings to determine their relevance to medication abortion. To be eligible for inclusion, each resource must be focused on medication abortion services including clinical (guidelines, protocols, checklists, and factsheets) and nonclinical services (financial, legal, and social). Since our objective was to improve equitable access to abortion among underserved populations, medication abortion resources

focusing on Indigenous populations, racialized groups, and gender minorities were included. Resources were excluded if they were dated more than 20 years ago, not relevant to the Canadian context, beyond the scope of medication abortion, or not targeted to health care providers (ie, physicians, pharmacists, social workers, and midwives). Most papers were not relevant to the Canadian context because of the constant regulatory changes regarding the prescription, dispensing, and use of medication abortion in Canada since 2017 [15-17]. We also excluded resources that were exclusively focused on procedural abortion, contraception, or general reproductive health or illnesses. Procedural abortion resources were excluded because procedural abortion demands specialized training, and we considered a web platform as an inappropriate training source for those seeking to provide this skilled task. After screening the resources against the inclusion and exclusion criteria, 124 resources were excluded, leaving us with 37 resources for a detailed expert review. Figure 2 shows the data extraction and review process.

Figure 2. Resource extraction and screening process.



Content Review

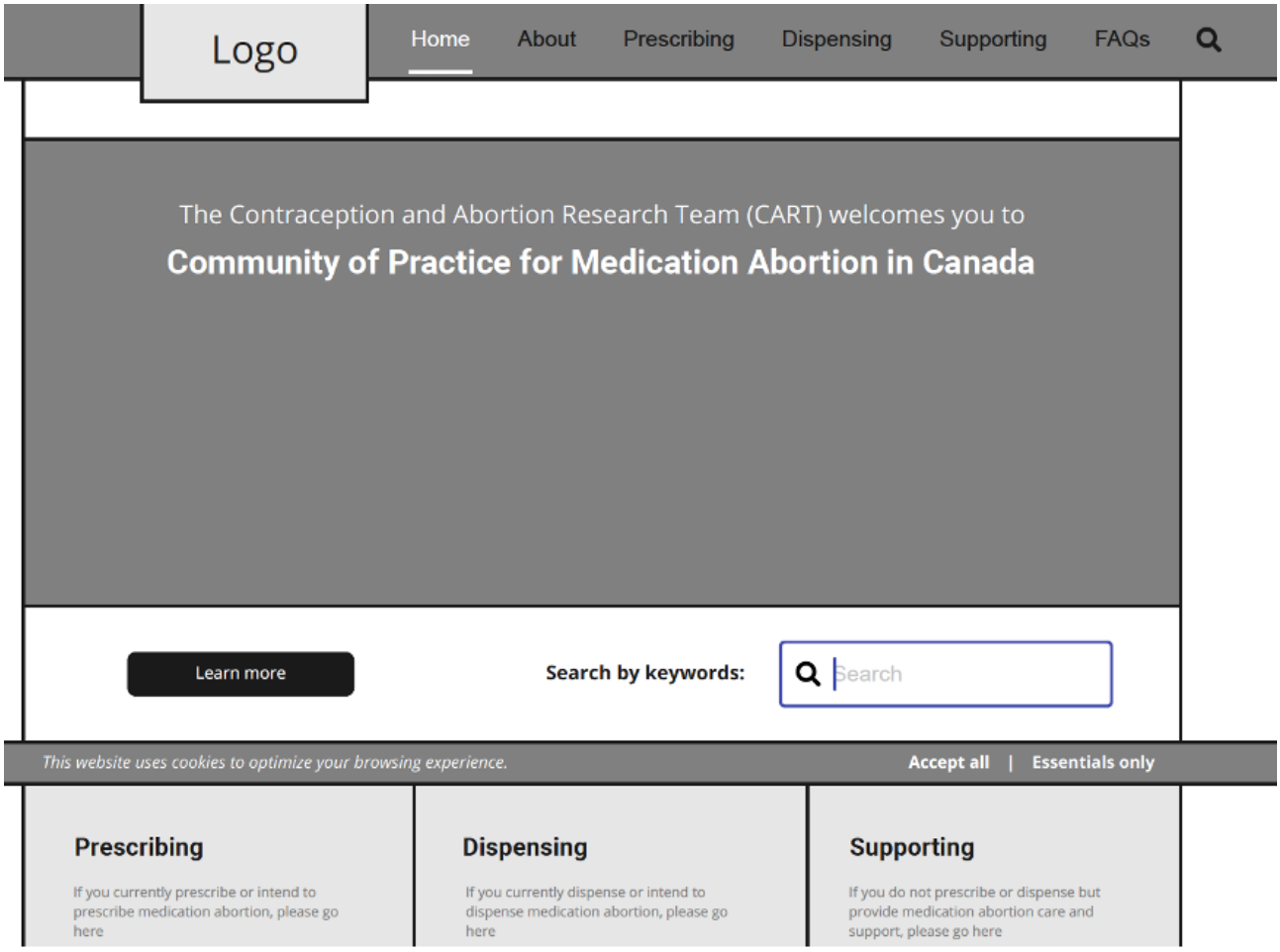
Even though the resources were extracted from credible sources, medication abortion regulations and professional practice guidelines in Canada have evolved over time, and with these changes, some of the resources might have become outdated. To ensure our website contains up-to-date resources and is in line with current practices, we engaged 10 health care and allied helping professionals providing abortion services to review the 37 resources for accuracy, currency, and relevance for providing medication abortion, particularly for equity-deserving populations. These health care professionals include physicians, nurse practitioners, pharmacists, and midwives. We asked each health care professional to review each document and respond to the following questions: (1) How do you think this resource will be relevant or useful for you to initiate or strengthen your medication abortion services and why? (2) Please indicate specific parts that you feel are most useful for your practice. (3) Please indicate specific parts that you think are out of date, not useful, or contradict your current practice.

All reviewers provided their feedback within 1 month, and all feedback was compiled into a single document. Following the review, we then engaged 2 experts in abortion care (a family physician and an obstetrician and gynecologist) to discuss and reconcile any differences or contradictory comments by the reviewers. The 2 expert reviewers resolved any differences and wrote a master document containing up-to-date resources on current medication abortion practices. The master file was organized into three main sections including (1) prescriber resources (how medication abortion works, preabortion medication evaluation, postabortion assessment, web-based and hybrid care, physician billing information, regulations, and inclusivity toolkit), (2) dispenser resources (coverage information, patient communication guidelines, checklist for dispensing abortion, and pre- and postabortion assessment), and (3) supporting resources (patient counseling before and after abortion and social worker support resources). This master document was used to inform the content of the website.

Phase 2: Development of Wireframes

Using a web-based design tool known as Figma (Figma Inc), we created wireframes of the website. A wireframe is a visual guide that shows a skeletal framework of a web interface [18]. The wireframe was designed to follow the format of the resources on the master file. The framework depicted the page layout and arrangement of content on the website including interface elements, navigational features, the range of functions to include, and the relative prioritization of content. This wireframe was reviewed by 3 experts in abortion care, and revisions were made regarding space allocation and content prioritization for each of the resource sections outlined in the master file. Following this review, the skeletal framework was then populated with the content from the master file. Figure 3 shows the wireframe of the website.

Figure 3. A screenshot of the wireframe.



Phase 3: Adapting the Wireframes to the Needs of Underserved Populations

After populating the skeletal framework with the revised content, we then conducted a focus group discussion with 14 health care professionals who currently provide, or intend to provide, medication abortion in Canada. The essence of the focus group was to understand how the wireframe can be adapted to equip health care professionals to provide inclusive, gender-affirming, and culturally safe abortions to underserved populations in Canada. We also wanted to identify potential areas for improvement before moving on to developing a functioning website.

Before the start of each focus group, we distributed a visual layout of the website containing relevant content to all participants. The participants were asked to review the visual layouts ahead of the web-based meetings. During the focus group discussions, participants were asked about (1) how the

information on the wireframe could be adapted to support them in providing evidence-based, culturally safe, and accessible abortion services to HPSM populations and (2) any additional features, functionalities, and considerations for ensuring confidentiality, appropriateness, culturally affirmativeness, and supportive resources for improving access and quality of abortion seeking experience for HPSM populations.

The focus groups were led by author AFA, with assistance from RC. Each session lasted approximately 1 hour, and the data were transcribed verbatim. The data from this focus group were analyzed thematically, and the findings (reported under the Results section) were used to modify the Figma designs before moving on to developing a functional product. The findings from the focus groups helped to minimize the design flaws and reduce the possibilities for major changes once a product was eventually developed.

Phase 4: Development of a Functional Website

After making changes to the Figma design, we engaged our software engineers to translate the Figma design into a functioning website. To further explain the Figma design to the development team, we wrote a Systems Requirements Specification Document to further explain the Figma design to the development team. This document included a breakdown of the Figma designs into various components and specifying each design feature as contained in the guide [19]. Figure 4 shows the interface of the functional website. To make sure that the functional product reflects the Figma design and meets the knowledge needs of the end-user clinicians, the development team sought ongoing feedback from our team as well as from the health care professionals who reviewed the resources. We held biweekly consultation meetings with the software development team starting from August 2023 until we had a functional website in February 2024. In compliance with the user-centered design process, we evaluated the functional website with end-user clinicians.

Figure 4. The home page of the functional website.

Phase 5: User Evaluation

We conducted a user evaluation of the website using a think-aloud observation method with 26 health care professionals across Canada. A think-aloud observation is a user evaluation method, in which participants verbalize their thought processes as they interact with a system [20]. The purpose of the user evaluation was to determine the relevance and usefulness of the website in providing abortion services to equity-deserving and HSPM populations, understand the ease of use of the website, detect any design flaws, and make final changes before the website is launched. During the think-aloud observation, we provided participants with 5 tasks related to information retrieval on our website. As participants navigated the website, they verbalized their thought processes by commenting on aspects of the website that they found confusing, difficult to navigate, or irrelevant to their purpose. The think-aloud data were analyzed using simple content analysis. The findings (reported under the Results section) were used to revise the website.

Results

Overview

In this section, we focus on the findings from the focus group discussions of the wireframes as well as the user evaluation of the functional website. Overall, the health professionals who took part in the focus group discussions and the end-user evaluation acknowledged that our web platform was an important educational resource that would equip them with the necessary knowledge to provide abortion care. They also provided recommendations on how this web-based platform could be adapted to support them in providing inclusive abortion care to underserved populations.

Findings From the Focus Group Discussion

Overview

The participants expressed their satisfaction with the layout and the content displayed on the wireframes. They indicated that the various clinical guidelines, factsheets, and checklists on medication abortion would help demystify the myths and misconceptions associated with providing abortion. The usefulness of the resources in demystifying abortion provision was a very important feature considering participants' self-described limited understanding of medication abortion. While acknowledging the importance and potential usefulness of the wireframes, participants noted some limitations on the interfaces of the wireframes and suggested ways in which they could be adapted to improve the capacity of health care professionals to provide abortion services to equity-deserving and HPSM populations. These recommendations fell into 4 thematic areas.

Inclusive and Culturally Sensitive Content

Participants expressed the need for the website to contain inclusive and culturally sensitive language that has the least chance of causing harm to underserved populations including Indigenous, trans, and nonbinary people who seek abortion. To enable health professionals to provide inclusive abortion services to trans people, the participants suggested that the website should contain gender-neutral terms as well as handouts and factsheets on abortion and postabortion care in multiple languages. A physician who provides abortion stated:

> *A lot of my patients can speak English but there are instances [where] I get patients that cannot speak English properly. In such cases, it will be good to have some resources in their language that I can give out to them.*

Some of them also suggested having a print-out function to provide materials for patients who do not speak English. The participants also expressed the need for low bandwidth connectivity to enable health care providers in rural areas with poor internet access to access resources.

Indigenous and Gender-Diverse Resources

The participants also called for the inclusion of Indigenous and 2SLGBTQI+ resources on the website because they indicated that one of the groups facing the most barriers to accessing health services includes Indigenous groups and the 2SLGBTQI+ populations. This could be seen in one of the statements given by one of the participants:

> *it isn't just about understanding the clinical guidelines and factsheets. But we also need to understand the other aspects of Indigenous cultural safety when providing abortion to Indigenous patients like the relationship between abortion care and legacies of forced apprehension/the millennial scoop, information on the knowledge and risk of reproductive coercion (either for or against abortion) etc.*

These calls were particularly made by participants who provide services to Indigenous communities. Other participants indicated a need for a learning hub or short courses on Indigenous cultural safety, gender-affirming, and trauma-informed abortion care—arguing that these additional knowledge sources will help them to provide services to abortion seekers who experienced historical and intergenerational traumas.

Providing Resources for Noninsured Clients

The participants also indicated how the website could be a "one-stop shop" of information if they included information on how to support noninsured clients. This recommendation was made by a midwife who revealed the challenges of providing abortion services for noninsured clients such as undocumented migrants:

> *More than half of my clients were unhoused, or street-involved or living in extreme poverty, or they were refugees, recently landed immigrants, or people of color and trans communities. You know there is a significant intersection between such populations and those living in poverty so there needs to be some resources around how to make sure that these populations are also getting access to abortion services.*

Protecting the Privacy of Website Users and Abortion Seekers

The participants acknowledged the stigmatizing nature of abortion and the need to protect people's privacy. Thus, they suggested including a built-in exit button that would enable users to immediately exit the website, particularly in public places. They also indicated the need for the website to contain geolocation features that can map out nearby abortion providers and pharmacies outside of people's communities. One participant stated:

> *I provide services in a local community, and some don't feel comfortable seeking abortion services from me because I may probably know them. When this happens, it's difficult to refer because they won't even come to you. So, it will be good to have a feature on this platform where patients themselves can search for nearby abortion providers that they can approach with comfort.*

According to the participants, this will make visible nearby abortion providers that could serve as alternative sources for patients who may be reluctant to seek abortion services within

their (particularly very small) communities due to privacy concerns.

Findings From the User Evaluation

Overview

Iterative feedback from the focus group informed changes to the website, as participants serially identified gaps and problems to be addressed before it was launched. Some health care professionals indicated that the functional website felt too medical and was insufficiently focused on the social aspects of abortion. They perceived our banner image on the home page as overrepresenting physicians and not sufficiently acknowledging the role of midwives and supporting professionals. They also made recommendations on how the website could be adapted for navigability and to touch on the stories of underserved populations. The recommendations from the user evaluation fell into 3 thematic areas.

Portraying a Positive Image of Abortion via Visual Effects

Some health care professionals commented that the initial wireframes lacked a human touch and thus suggested adding more vibrant and bright colors that reflect the identities of diverse groups of health professionals and the diversity of patients. Participants felt that such adaptations could portray abortion as something that is considered normal for everyone to seek, normal for every health professional to provide or support, and to ensure the abortion-seeking experience is positive.

> *maybe some colors will be nice because I feel that when I am looking into abortion care, I just want to take this stigma away. So, having colors that are dark for me is like you are doing something bad, right? So maybe having some lighter colors and highlighting on socio-cultural aspects of abortion would be helpful. People need to feel that abortion is normal after going through this website.*

Connecting With Patients

Other health professionals explained that although the website targets health care professionals, it is ultimately about caring for people. They thus suggested that it would be great to have images of health care providers taking care of patients from diverse gender and racial backgrounds or better still include images of patients having children to remove the myth that people seeking abortion are irresponsible adults without children or families. One participant stated:

> *This website is about people getting care. So, we need to see how so many people who have pregnancy termination also have children already. So, it's not just like the unwed single woman who wants to terminate her pregnancy.*

The participants also expressed the need for the website to help health care professionals understand the patient populations most likely to seek abortion services from them. This way, the website would not only provide clinical information but also help providers to connect with the patients' stories.

Making Compromises Between Being Comprehensive and Practical

Many participants acknowledged the comprehensiveness of the website and were confident that they would be able to find relevant information should the need arise. They also acknowledged that being comprehensive might make the website a less suitable quick reference guide, particularly during clinician-patient encounters. This concern could be a challenge for high-volume providers and providers in rural and remote areas. One midwife stated:

> *the information that I'm seeing is actually really good. It's really good information, and I'm intrigued to read more about what's in here. But [it's] too much. It would be really good for like education and training, but not when you have a patient in front of you.*

While this theme was not specifically related to providing abortion to underserved populations, it highlighted the busy schedule of clinicians vis-à-vis looking for information in a dense and cluttered web platform.

Discussion

Principal Findings

In this study, we report on the cocreation of an open-access web-based abortion platform and highlight considerations for promoting access to abortion for underserved populations. This study followed an intersectionality study we conducted with end-user clinicians before embarking on the development of the website [21]. Our study incorporated rigorous engagement of end-user clinicians, and our user evaluation data demonstrated an accessible, acceptable website that could better equip health and allied helping professionals to provide equitable, culturally safe, and gender-affirming resources to equity-deserving and HSPM populations. It is important to note that many health professionals in Canada, including nurse practitioners and midwives, had not received any formal training on abortion as part of their education, leading to limited knowledge and general uncertainties in providing abortion [22]. The limited knowledge may lead to some hesitation in providing abortion, particularly for patients from underserved populations who present with complex health challenges. We believe that this mentorship platform will equip health professionals with the necessary knowledge to provide abortion to patients who face intersecting barriers to access services while also enhancing timely decision-making for health professionals in rural and remote areas. For a time-sensitive procedure such as abortion, expeditious decision-making is crucial to high-quality care for women and pregnancy-capable people who request abortion services.

In addition to information dissemination and helping with decision-making, an open-access abortion web platform may also serve to counter abortion-related misinformation and disinformation that are often targeted at health care professionals. This is important at a time when the proliferation of disinformation, together with the stigma and belief-based refusal to provide legal and appropriate abortion care, creates interpersonal tensions that make it difficult for some health

professionals to meet the reproductive needs of their patients [23,24]. Health professionals may further be disadvantaged when there are frequent changes in medication abortion guidelines and regulations [17]. These changes, together with the need for quality improvement, highlight the importance of such a web-based platform for updating the competencies of health care professionals who provide or intend to provide abortion. This is more relevant, as there is an increasing call for the timely integration of research evidence in routine clinical practice [25].

Our findings suggest that adopting a user-centered design in this project resulted in a platform that was largely considered by potential end-user clinicians as inclusive. While engaging stakeholders from diverse clinicians and organizations was quite challenging, we acknowledged that the process was essential to fulfill the principles of user-centered design. Our team believes that the active engagement of clinicians, health professional organizations, and abortion advocacy groups in cocreating this web-based platform enhanced the appropriateness of the content while facilitating clinician recruitment, dissemination, and subsequent uptake of the resource among end-user health care professionals.

It is also important to note that few digital health projects are designed from gender equity and Indigenous rights perspective [26]. In the context of abortion-related websites, intersectionality is particularly important, as it helps illuminate ways in which the website could be tailored to the needs of the population subgroups facing the most barriers in access to abortion (racialized groups, migrants, people with disabilities, homeless and underhoused people, sex workers, 2SLGBTQI+ people, Indigenous people, and youths) [21,27]. Indeed, available evidence suggests that such minority groups are overrepresented among those with unintended pregnancies and particularly among those seeking abortion [28-30]. Therefore, engaging with a group of health professionals who serve diverse populations in developing and evaluating our web platform made the website unique, as it cedes all decisions on the content and structure in the hands of the end-user clinicians who may have first-hand experience in providing services to underserved populations. This approach to developing our website supports the vision of “design justice” where the community owns the design artifacts [31]. By cocreating with health care professionals, we were able to generate website content that illuminates ways health care professionals can provide abortion services to underserved populations who present with complex, historical, and intersecting health challenges. This approach is more likely to lead to a seamless translation of the web platform into practice since the clinicians and partner organizations were more or less coowners of the end product.

Creating web platforms that are inclusive of equity-deserving and HSPM populations requires more than just technically sophisticated algorithms but a deep understanding of users’ needs and specific considerations that promote inequities in access to health care [32]. The inclusive design recommendations from the participants in this study were aimed at equipping health care professionals with the necessary tools to address inequities in access to abortion care. The integration of low-bandwidth connectivity for instance would improve accessibility of the resources to health professionals in rural and remote areas with low internet connectivity. Improved web access could help improve abortion access in rural and remote areas where medication abortion might be the only, or most accessible, option [33]. Furthermore, the recommendation for including geolocation services could help address abortion-related stigma and provide better privacy for abortion seekers in communities with conservative values around abortion. While the argument for geolocation features was seen as a way of helping patients easily locate abortion providers and pharmacists, some were concerned that geolocation services could be used by antichoice groups to target abortion providers. Additional security measures may be needed if geolocation features are to be implemented on abortion-related web platforms.

Limitations

Our study was not without limitations. Even though our objective was to design a web platform that supports health care professionals to provide equitable access to abortion for underserved populations, we did not rigorously engage abortion seekers the same way we engaged with clinicians. A future improvement would be to engage the patients’ voices in resource development to better reflect the peculiar needs of underserved populations. Our ultimate goal for this website is to empower health care professionals to provide equitable and inclusive abortion to underserved populations. However, we cannot ascertain if we have achieved that objective at the time of writing this paper. Future evaluation reports from users would tell us if the resources provided in this platform are indeed useful in providing abortion services to underserved populations. Furthermore, the participants made several recommendations in the focus group and the user evaluation that were not implemented mainly due to time and resource constraints. While we acknowledge this as a limitation, it is not uncommon to see such limitations in digital health projects, as it is practically impossible to implement all findings from user evaluation [34,35]. Finally, even though we planned to recruit health professionals who identify as or have expertise in providing abortion services to underserved populations (ie, racialized people, migrants, people with disabilities, homeless and underhoused people, sex workers, 2SLGBTQI+ people, and youths as well as those living in or providing abortion services in rural or remote areas), it was difficult to find health professionals who identify as or has expertise in all these diverse populations. Therefore, some phases were completed without adequate input from health professionals with expertise or experience in providing abortion services to underserved populations.

Conclusions

Drawing on a user-centered design approach, we cocreated a gender-affirming and culturally appropriate open-access mentorship platform with and for health care professionals to support the equitable provision of abortion in Canada. We believe that a platform of this nature would increase medication abortion awareness among health care professionals and destigmatize abortion care while clarifying the myths and misconceptions that often create tensions and foment general

uncertainties in initiating and sustaining abortion care. With underserved populations more likely to have unintended pregnancies than the general population [36,37], a website that specifically seeks to improve the competencies of health care professionals in providing abortion services to this population is important and timely. We believe that this platform has the potential to enhance the knowledge and expertise of health care professionals, particularly those in rural and remote areas on how to initiate and sustain abortion care for HPSM populations in Canada who face the most barriers in access to health services. This platform is also expected to provide the necessary resources needed to support health care professionals in providing evidence-based and culturally safe medication abortion for health care professionals in rural and remote areas.

Acknowledgments

The funding for this project was received from the Health Canada Sexual and Reproductive Health Fund under the Health Care Policy and Strategies Program. The authors would also like to acknowledge the Health Professional Association (ie, Society of Obstetricians and Gynaecologists of Canada, Canadian Pharmacy Association, Canadian Association of Midwives, and Canadian Association of Schools of Nursing), advocacy groups (Action Canada for Sexual Health and Rights and National Abortion Federation), and health care professional participants who provided diligent reviews of the resources.

Conflicts of Interest

None declared.

References

1. Sexual and reproductive health and rights. Office of the High Commissioner for Human Rights. URL: https://www.ohchr.org/en/women/sexual-and-reproductive-health-and-rights [accessed 2024-06-05]
2. Black AY, Guilbert E, Hassan F, et al. The cost of unintended pregnancies in Canada: estimating direct cost, role of imperfect adherence, and the potential impact of increased use of long-acting reversible contraceptives. J Obstet Gynaecol Can 2015 Dec;37(12):1086-1097. [doi: 10.1016/s1701-2163(16)30074-3] [Medline: 26637081]
3. Dunn S, Cook R. Medical abortion in Canada: behind the times. CMAJ 2014 Jan 7;186(1):13-14. [doi: 10.1503/cmaj.131320] [Medline: 24277708]
4. Schummers L, Norman WV. Abortion services in Canada: access and safety. CMAJ 2019 May 13;191(19):E517-E518. [doi: 10.1503/cmaj.190477] [Medline: 32392474]
5. Monchalin R, Jubinville D, Pérez Piñán AV, et al. “I would love for there not to be so many hoops … ”: recommendations to improve abortion service access and experiences made by Indigenous women and 2SLGTBQIA+ people in Canada. Sex Reprod Health Matters 2023 Dec;31(1):2247667. [doi: 10.1080/26410397.2023.2247667] [Medline: 37799036]
6. Monchalin R, Pérez Piñán AV, Wells M, et al. A qualitative study exploring access barriers to abortion services among Indigenous Peoples in Canada. Contraception 2023 Aug;124:110056. [doi: 10.1016/j.contraception.2023.110056] [Medline: 37116606]
7. Sethna C, Doull M. Spatial disparities and travel to freestanding abortion clinics in Canada. Womens Stud Int Forum 2013 May;38:52-62. [doi: 10.1016/j.wsif.2013.02.001]
8. Ennis M, Renner R, Olure B, Begun S, Norman WV, Munro S. Provision of care to diverse populations: results from the 2019 Canadian Abortion Provider Survey. BMJ Sex Reprod Health 2024 Oct 15;50(4):308-309. [doi: 10.1136/bmjsrh-2023-202175] [Medline: 38729768]
9. Decolonizing reproductive justice—indigenous voices. Abortion Rights Coalition of Canada. URL: https://www.arcc-cdac.ca/activist-workshops/decolonizing-rj-indigenous-voices/ [accessed 2024-06-06]
10. Monchalin R. Novel coronavirus, access to abortion services, and bridging Western and Indigenous knowledges in a postpandemic world. Womens Health Issues 2021;31(1):5-8. [doi: 10.1016/j.whi.2020.10.004] [Medline: 33371932]
11. Georgsson S, Carlsson T. Readability of web-based sources about induced abortion: a cross-sectional study. BMC Med Inform Decis Mak 2020 Jun 5;20(1):102. [doi: 10.1186/s12911-020-01132-y] [Medline: 32503524]
12. Anderson NS, Norman DA, Draper SW. User centered system design: new perspectives on human-computer interaction. Am J Psychol 1988 ;101(1):148. [doi: 10.2307/1422802]
13. CAPS CPCA Forum. URL: https://caps-cpca.ubc.ca/index.php?title=Main_Page [accessed 2023-09-07]
14. AusCAPPS | Medcast. URL: https://medcast.com.au/communities/auscapps [accessed 2024-06-06]
15. Munro S, Wahl K, Soon JA, et al. Pharmacist dispensing of the abortion pill in Canada: diffusion of innovation meets integrated knowledge translation. Implement Sci 2021 Aug 3;16(1):76. [doi: 10.1186/s13012-021-01144-w] [Medline: 34344393]
16. Stirling-Cameron E, Carson A, Abdulai AF, et al. Nurse practitioner medication abortion providers in Canada: results from a national survey. BMJ Sex Reprod Health 2024 Dec 30. [doi: 10.1136/bmjsrh-2024-202379] [Medline: 39414352]

17. Munro S, Guilbert E, Wagner MS, et al. Perspectives among Canadian physicians on factors influencing implementation of mifepristone medical abortion: a national qualitative study. Ann Fam Med 2020 Sep;18(5):413-421. [doi: 10.1370/afm.2562] [Medline: 32928757]
18. What is wireframing? The Interaction Design Foundation. URL: https://www.interaction-design.org/literature/topics/wireframe?srsltid=AfmBOoo2DUHZGcYHDMRnK52HW6gksg6tgPoUvH6MLDMpQFe3BS_-q1xy [accessed 2023-09-13]
19. How to write system requirement (SRS) documents. Visure Solutions. URL: https://visuresolutions.com/requirements-management-traceability-guide/how-write-system-requirement-documents/ [accessed 2024-01-24]
20. Boren MT, Ramey J. Thinking aloud: reconciling theory and practice. IEEE Trans Profess Commun 2000;43(3):261-278. [doi: 10.1109/47.867942]
21. Abdulai AF, Czerniak E, Duong C, et al. Intersectionality in developing a virtual community of practice platform on abortion. Stud Health Technol Inform 2024 Jul 24;315:392-397. [doi: 10.3233/SHTI240176] [Medline: 39049289]
22. Paynter M, Norman WV, Martin-Misener R. Nurses are key members of the abortion care team: why aren't schools of nursing teaching abortion care? CJCND 2019 Dec 10;1(2):17-29. [doi: 10.25071/2291-5796.30]
23. Pagoto SL, Palmer L, Horwitz-Willis N. The next infodemic: abortion misinformation. J Med Internet Res 2023 May 4;25:e42582. [doi: 10.2196/42582] [Medline: 37140975]
24. Ennis M, Renner RM, Olure B, et al. Experience of stigma and harassment among respondents to the 2019 Canadian abortion provider survey. Contraception 2023 Aug;124:110083. [doi: 10.1016/j.contraception.2023.110083] [Medline: 37263373]
25. Johnston HB, Footman K, Ali MM, et al. Measuring availability of and facility readiness to deliver comprehensive abortion care: experiences and lessons learnt from integrating abortion into WHO's health facility assessments. BMJ Glob Health 2024 Aug 9;8(Suppl 4):e015097. [doi: 10.1136/bmjgh-2024-015097] [Medline: 39122445]
26. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021 Aug;3(8):e526-e533. [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
27. Price K. It's not just about abortion: incorporating intersectionality in research about women of color and reproduction. Womens Health Issues 2011;21(3 Suppl):S55-S57. [doi: 10.1016/j.whi.2011.02.003] [Medline: 21530842]
28. Mark A, Foster AM, Aguilar G, et al. Access for all: research at the National Abortion Federation's 47th Annual Meeting. Contraception 2024 Jun;134:110402. [doi: 10.1016/j.contraception.2024.110402]
29. Foster AM, Darney BG, Nguyen BT, Russo J, Arora KS, Westhoff C. Advancing diversity, equity, inclusion, and anti-racism in contraception: a statement of commitment from the Editorial Board. Contraception 2023 Jul;123:110066. [doi: 10.1016/j.contraception.2023.110066] [Medline: 37207728]
30. Hulme J, Dunn S, Guilbert E, Soon J, Norman W. Barriers and facilitators to family planning access in Canada. Healthc Policy 2015 Feb;10(3):48-63. [Medline: 25947033]
31. Udoewa V. An introduction to radical participatory design: decolonising participatory design processes. Des Sci 2022;8:e31. [doi: 10.1017/dsj.2022.24]
32. Inclusive technology: design for underserved and marginalized communities. The Department of Computer Science at the University of Chicago. URL: https://www.classes.cs.uchicago.edu/archive/2020/winter/20370-1/ [accessed 2024-04-27]
33. Sandoval S, Chen G, Rafie S, Rabin B, Mody S, Averbach S. A qualitative analysis of pharmacists' attitudes towards provision of medication abortion. BMC Health Serv Res 2023 May 30;23(1):557. [doi: 10.1186/s12913-023-09543-z] [Medline: 37254171]
34. Chapter 10: MoSCoW prioritisation. Agile Business Consortium Limited. URL: https://www.agilebusiness.org/dsdm-project-framework/moscow-prioririsation.html [accessed 2024-01-06]
35. Lehtola L, Kauppinen M, Kujala S. Requirements prioritization challenges in practice. In: Bomarius F, Iida H, editors. Product Focused Software Process Improvement: Springer; 2004:497-508. [doi: 10.1007/978-3-540-24659-6_36]
36. Reynolds CA, Beccia A, Charlton BM. Multiple marginalisation and unintended pregnancy among racial/ethnic and sexual minority college women. Paediatr Perinat Epidemiol 2021 Jul;35(4):493-500. [doi: 10.1111/ppe.12744] [Medline: 33345309]
37. Oulman E, Kim THM, Yunis K, Tamim H. Prevalence and predictors of unintended pregnancy among women: an analysis of the Canadian Maternity Experiences Survey. BMC Pregnancy Childbirth 2015 Oct 13;15(1):260. [doi: 10.1186/s12884-015-0663-4] [Medline: 26462914]

Abbreviations

2SLGBTQI+: two-spirit, lesbian, gay, bisexual, transgender, queer, intersex, and gender-diverse
HPSM: historically, persistently, or systemically marginalized

Edited by A Kushniruk; submitted 19.06.24; peer-reviewed by L Patchen, WR Dullius; revised version received 31.12.24; accepted 05.01.25; published 19.02.25.

Please cite as:
Abdulai AF, Duong C, Stroulia E, Czerniak E, Chiu R, Mehta A, Koike K, Norman WV
A Health Professional Mentorship Platform to Improve Equitable Access to Abortion: Development, Usability, and Content Evaluation
JMIR Hum Factors 2025;12:e63364
URL: https://humanfactors.jmir.org/2025/1/e63364
doi:10.2196/63364

Developing an App for Real-Time Daily Life Observations in a Nursing Home Setting: Qualitative User-Centered Co-Design Approach

Coen Hacking, MSci; Bram de Boer, PhD; Hilde Verbeek, PhD; Jan Hamers, PhD; Sil Aarts, PhD

Department of Health Services Research, Maastricht University, Duboisdomein 30, Maastricht, the Netherlands

Corresponding Author:
Bram de Boer, PhD
Department of Health Services Research, Maastricht University, Duboisdomein 30, Maastricht, the Netherlands

Abstract

Background: Assessing the daily lives of older adults, including their activities, social interactions, and well-being is essential, particularly in nursing homes, as it gains insights into their quality of life. Methods such as the Microsoft Excel-based Maastricht Electronic Daily Life Observation (MEDLO) tool are time-consuming and require extensive manual input, making them difficult to use.

Objective: This study aimed to develop an app-based version of the MEDLO using a user-centered design (UCD) and co-design approach to enhance efficiency and usability. We looked to actively involve researchers and care professionals who have used the MEDLO before, throughout the development process.

Methods: Participants included a diverse group of researchers and care professionals experienced in using the MEDLO tool. The UCD approach involved multiple iterative phases including semistructured interviews, user research sessions, and application development. Data were analyzed using a qualitative (thematic) approach of UCD and user research sessions. The app, which was preferred to the traditional Excel-based MEDLO, underwent multiple iterations. This method primed the continuous iterative development of the app, aimed for a minimum viable product (MVP).

Results: This study included 14 participants, primarily female, from diverse professional backgrounds. Their feedback highlighted the need for efficiency improvements in tool preparation and data management. Key improvements included automated data handling, an intuitive tablet interface, and functionalities such as randomization and offline data syncing.

Conclusions: The iterative development process led to an app that aligns with end-user needs, indicating potential for improved usability. Early and continuous user involvement was key in enhancing the application's usability, demonstrating the importance of user feedback in the development process.

(JMIR Hum Factors 2025;12:e57911) doi:10.2196/57911

KEYWORDS

co-design; user-centered design; app development; nursing home; user-centered; design; efficiency; usability; tablet; mobile phone

Introduction

Assessing the daily lives of older adults, particularly in the context of nursing homes, is essential for gaining insights into quality of life [1,2]. It allows health care providers to detect and address issues related to their physical health (eg, mobility limitations, pain), mental well-being (eg, mood disturbances, cognitive decline), and social interactions (eg, isolation, engagement in activities) in a timely manner [1,3,4]. This information can then be used to personalize care plans, thereby potentially improving health outcomes and enhancing the overall quality of life of older adults [5,6].

Ecological momentary assessments (EMAs) have been developed to facilitate this process by providing real-time, on-site evaluations of an individual's well-being [5,7,8]. EMA involves collecting data on individuals' behaviors and experiences in their natural environments, which can then be used to identify patterns and inform interventions [7]. However, existing EMA tools are mainly used in research settings, and their implementation in clinical practice has been challenging due to the time-consuming nature of data collection and the complexity of the tools [9,10]. The Maastricht Electronic Daily Life Observation tool (MEDLO-tool) is designed to assess the daily lives of nursing home residents using EMA methodologies [1,4]. The tool captures real-time information across several key dimensions of daily life, including activity levels (eg, participation in communal or solitary activities), agitation, mood states, and interactions with staff, fellow residents, and visitors [1,4,11]. By systematically observing and recording these

aspects, the MEDLO-tool provides a nuanced view of residents' experiences, which can inform personalized care strategies [4]. However, the MEDLO-tool relies on Excel templates that require significant manual input and are not user-friendly for care professionals [1,4]. The complexity of the Excel-based system poses significant challenges, including data entry errors, time inefficiency, and barriers to widespread adoption in clinical settings [12,13]. This complexity can be particularly problematic in nursing homes, where staff may have limited time and technological proficiency [13].

Developing a mobile application for the MEDLO-tool could address these challenges by automating data collection processes, reducing manual input, and providing an intuitive interface for users [14,15]. An app could streamline observations, enable real-time data analysis, and facilitate immediate feedback to care providers, thereby enhancing the tool's use in clinical practice. Furthermore, considering the unique needs of residents with dementia, it is important that such an app is designed to account for cognitive impairments and communication difficulties [16,17]. The development of such an app requires a thorough understanding of the needs and preferences of the end-users. A user-centered design (UCD) approach is vital for developing software that truly meets users' needs [14,18]. UCD emphasizes involving users throughout the design process to ensure that the product aligns with their requirements and preferences [19]. This involves iterative cycles of user need assessment, design, prototyping, and testing [20].

Co-design, a component of UCD, involves direct collaboration between users and designers, requiring the active participation of users in generating ideas, making decisions, and solving problems [21,22]. This not only ensures the software's functionality but also its usability [21].

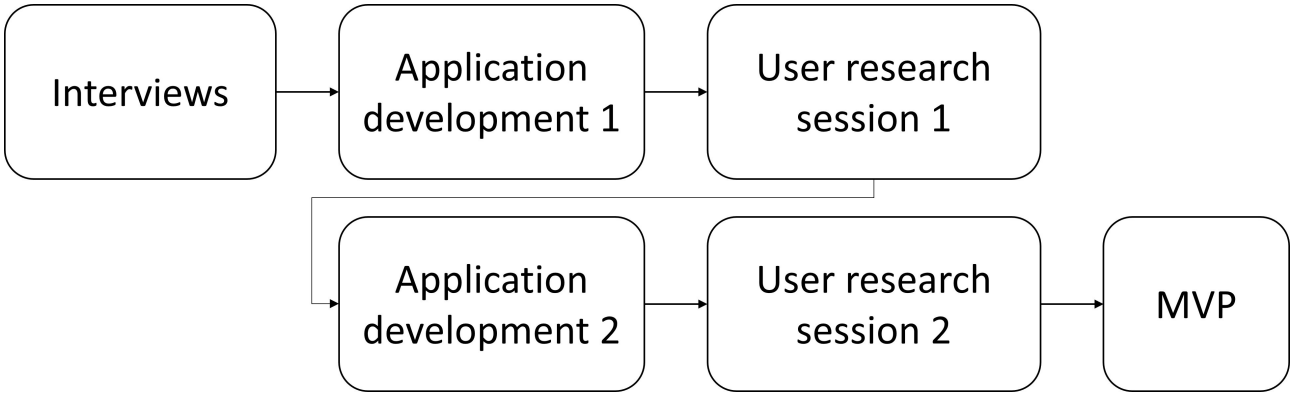
Potential barriers to implementing such technology include technological literacy among staff, user engagement, and ensuring data privacy and security [13,23]. By actively involving care professionals and stakeholders in the design process, we aim to create a tool that is both functional and user-friendly, facilitating its adoption in clinical practice and ultimately improving the quality of care for residents with dementia. Therefore, this study is aimed at developing an app for measuring the daily life of residents with dementia living in nursing homes, using a user-centered and co-design approach.

Methods

Study Design

This methodological study aimed to develop a mobile application version of the MEDLO-tool for use in long-term care facilities, using a UCD approach [24-26]. The study unfolded in multiple iterative phases, including semistructured interviews, user research sessions, application development, and comprehensive data analysis, as illustrated in Figure 1. The project culminated in the achievement of a minimum viable product (MVP).

Figure 1. The iterative process of app development. MVP: minimum viable product.



Participants

Participants were selected through purposive sampling to include individuals with experience using the MEDLO-tool. The sample consisted of researchers and health care professionals affiliated with Maastricht University and local nursing homes. Inclusion criteria were previous use of the MEDLO-tool and willingness to participate in the study. A total of 14 participants were recruited, comprising research assistants, PhD candidates, postdoctoral researchers, senior researchers, occupational therapists, research coordinators, and associate professors.

Data Collection

Data collection involved semistructured interviews and user research sessions conducted between January and June 2023.

Interviews

Semistructured interviews were guided by a comprehensive list of topics derived from available literature and consultation with the researchers who developed the original MEDLO-tool [1,4]. The interview guide can be found in Multimedia Appendix 1. The interviews aimed to gather insights into participants' experiences with the Excel-based MEDLO-tool, usability issues, functionality requirements, data analysis techniques, and general impressions [27]. Each interview lasted approximately 45 to 60 minutes and was audio-recorded with participants' consent. The audio was transcribed verbatim for analysis. The interview guide is provided in the supplementary materials.

The interview transcriptions were subjected to a 6-phase thematic analysis process, aligning with the methodologies supported by previous research [28-30]. The inductive analysis was conducted by one researcher. A second researcher reviewed

the coding to enhance trustworthiness. This comprehensive process involved familiarization with the data, generating initial codes, identifying themes, reviewing themes, defining and naming themes, and ultimately, reporting the results. The resulting report summarized the key findings, interpretations, and implications, to provide valuable insights for the next stages of the development of the app.

User Research Sessions

A total of 2 user research sessions were conducted monthly to involve participants in the iterative development process. The session guide can be found in Multimedia Appendix 2. Each session lasted approximately 2 hours and was held at Maastricht University. Participants interacted with app prototypes using tablets and smartphones provided by the research team. The sessions included guided tasks (eg, completing an observation using the app), open-ended discussions, and interactive feedback activities facilitated by researchers using prototypes and interactive demos. The structure of these sessions was informed by previous co-design methodologies [1,3,31].

Data Analysis

Qualitative Analysis of Interviews

Interview recordings were transcribed verbatim and analyzed using thematic analysis following Braun and Clarke's 6-phase approach [28]. A total of 2 researchers independently coded the transcripts to enhance reliability. Initial codes were generated and organized into potential themes. Discrepancies in coding were discussed and resolved through consensus meetings. Themes were reviewed and refined to ensure they accurately represented the data. The final themes captured key insights into the usability and functionality of the MEDLO-tool, informing the app development. The overall process of thematic analysis was guided by earlier works in this field [29,30]. The session guide is provided in Multimedia Appendix 1.

User Research Sessions

Each user research session included guided tasks (eg, completing an observation using the app), open-ended discussions, and interactive feedback activities facilitated by researchers using mock-ups and interactive demos. Feedback from user research sessions was documented through field notes and audio recordings. An inductive content analysis was conducted by one researcher, who coded feedback and categorized issues into 4 themes: layout, functionality, errors, and appearance. A second researcher reviewed the coding to enhance trustworthiness. This categorization allowed the development team to prioritize user concerns based on frequency and severity, ensuring a user-centric approach to app refinement.

Application Development

The application was developed using ".NET MAUI" (ie, a framework designed by Microsoft) for cross-platform compatibility on iOS and Android devices. The server-side (ie, where the data are sent to) used "ASP.NET Core" (ie, a framework designed by Microsoft) for a maintainable and performant application. The development process was iterative, with weekly meetings between developers and researchers to integrate user feedback from the interviews and user research sessions.

Usability testing involved participants completing specific tasks using the app prototypes while researchers observed and noted any issues. Testing sessions were conducted in a controlled environment to simulate actual usage conditions. Feedback from these tests informed further refinements and addressed potential adoption or feasibility issues.

Ethical Considerations

This study did not require formal ethics approval under the Medical Research Involving Human Subjects Act (WMO). In the Netherlands, research that collects only anonymized or nonsensitive feedback—without requiring participants to undergo procedures that affect their physical or psychological integrity—is not considered WMO research and is therefore exempt from ethics review. The guidelines provided by the Central Committee on Research Involving Human Subjects clarify that retrospective or noninvasive studies using aggregated or anonymized data do not fall under the scope of the WMO [32]. The Maastricht University Medical Center+ Research Code also confirms that studies conducted with minimal risk, such as those analyzing collective feedback, do not require formal ethical review. All participant feedback was anonymized and is presented in aggregate form to ensure confidentiality [33].

Results

Sample

The study involved 14 participants with an average age of 36 (SD 10; median 39; range 24 - 57) years, and 93% (13/14) of them were female. The participants represented a diverse range of professions, including research assistants (n=3), PhD candidates (n=4), postdoctoral researchers (n=2), a senior researcher (n=1), an occupational therapist (n=1), a research coordinator (n=1), an associate professor (n=1), and an individual who was currently unemployed. Most participants (9/14, 64%) were affiliated with Maastricht University, while others worked at health care providers (4/14, 29%) or were unemployed (1/14, 7%). The average duration of employment at their respective institutions was 6 (SD 5; median 6; range 0.5 - 16) years.

All 14 participants had previous experience with the MEDLO-tool; 12 (86%) participants had used it directly in their research, while 2 (14%) participants had used it as inspiration for developing other tools different from MEDLO. All participants had also been in contact with the developers of the MEDLO-tool. Table 1 presents the demographic information of the participants in detail.

Table . Demographic information of the participants.

Characteristic		Values
Age (years), mean (SD)		
	24 - 57	36 (7.7)
Gender, n (%)		
	Female	13 (93)
	Male	1 (7)
Time at institution, mean (SD)		
	6 months to 16 years	6 (3.3)
Use of the tool, n (%)		
	Used in research	12 (86)
	Inspiration for another tool	2 (14)
Contact with developers, n (%)		
	All participants	14 (100)
Participants' affiliations, n (%)		
	University	9 (64)
	Health care provider	4 (29)
	Unemployed	1 (7)

Interviews

The participants described the process of using the MEDLO-tool in research projects as comprising 7 distinct phases: (1) acquiring informed consent, (2) preparing the Excel sheets before the observation, (3) familiarizing themselves with the faces of the residents who would be observed, (4) conducting the observations, (5) fixing issues with the data (such as missing values or inconsistencies), (6) analyzing the data, and (7) communicating the data back to the care organizations.

Participants emphasized that the preparation of the tool and the subsequent data cleaning were particularly time-consuming and would significantly benefit from automation. One participant stated, "Every time before we do a new observation, we have to look at the participants, randomize them several times, and input all of this into the Excel sheet." Another participant highlighted the challenges faced after data collection, "After the observation is done, I have to go through the whole document to check all the fields that I didn't have time to fill out or that I didn't know how to rate." These comments reflect the manual and labor-intensive nature of the existing process, which participants found cumbersome.

Participants reported that during observations, they used separate pieces of paper to write down descriptions of each resident's appearance because the observation periods were too short to input this information into the Excel sheet in real time. They suggested that having an additional field to securely write this information directly on the tablet would make the process more accessible. One participant explained, "We usually arrive at least half an hour earlier, so we can ask the nurses which participant corresponds to which name, and write down what they look like. We do that to find them back more easily during the observation." Another participant elaborated on the logistical challenges, "The data entry process on the tablet is intuitive, but we also have to keep track of the patient names and descriptions on paper, we carry the manual, and we use a timer on our phone to ensure that we don't spend too long on any resident. This is a lot to carry around." Regarding data reporting, participants indicated that the data reported back to the nursing homes primarily consisted of quantitative measures, which could potentially be automated through a dashboard (eg, Microsoft PowerBI). Participants expressed that automating this process would reduce the time between data collection and reporting, which currently could be several months. One participant noted, "When we want to report back the numbers to the care organization, we make a PowerPoint presentation that includes aggregated numbers. This usually takes place several months after the observations." Another participant suggested: "I think some of the numbers could be reported back automatically."

Overall, participants preferred using a tablet compared to using pen and paper, citing the intuitive data entry process. However, they also mentioned that a tool compatible with smartphones would be beneficial, especially for care professionals who might find tablets less convenient. One participant remarked, "I think the tool works well, but I'm not sure if it would be usable like this for nurses."

Initial User Research Session

Based on the information gathered during the interviews and the suggested points for improvement, an initial version of the application was created and provided to the participants for the first user research session. This first draft of the application consisted of an dashboard web application and a mobile app, both of which included the core components but had limited functionality. Multimedia Appendix 3 shows the app and dashboard as they were presented to the participants. Multimedia

Appendix 4 provides a detailed overview the feedback received during the initial user research session.

During the initial user research session, participants expressed largely positive sentiments towards the concept of the application and appreciated the efforts made in its development. One participant commented, "I know we have a lot of complaints about the app, but we really appreciate the work [the development team] has done already." However, participants also identified several areas for improvement. They expressed concerns regarding certain features of the app, such as the timer functionality, which lacked a reset option. Participants noted that in their current practice, they used interval timers on their phones to manage observation times for each resident, and the app's timer did not adequately support this need. One participant suggested, "What could be nice is some kind of interval timer. That's what we currently use on our phones." In addition, participants missed the comprehensive overview provided by the Excel-based MEDLO-tool, as the prototype app had been designed primarily for phone screens and did not offer the same level of data visibility. One participant observed, "Normally, you could easily scroll back in the overview." Furthermore, there were some unclear aspects of the app, such as certain functionalities being available only through the dashboard and not within the app itself. Participants found labels on buttons to be unclear, which led to confusion during navigation. They recommended improving the screen layout for better usability and enabling easy switching between participants by a list.

Participants raised concerns about data storage and security, even though these aspects had been addressed in the app's design. They questioned how the data was stored and whether it was encrypted. It was explained to them that the data was stored on the device (ie, phone or tablet) in encrypted form and could be synced to a secure server. Regarding randomization, participants asked how it would be conducted within the app and where they could access this feature. It was clarified that randomization was performed automatically upon creating a new observation group. A participant inquired, "How do I randomize the participants in the app?" To which the response was, "That is done automatically when an observation group is added."

Participants inquired about features that had not yet been implemented but were planned for future updates. For example, they asked if the app would work offline, which was an important consideration for use in environments with limited internet connectivity. They also asked whether the data could be exported to Excel for further analysis. These features were already part of the application's development roadmap but were not yet available in the user interface at the time of the session. One participant expressed concern, "I don't know if it's an issue when I lose connection, or whether I lose my data." Another participant discussed data export capabilities, "These are the data you see when you download the Excel file. With Excel, it looks a bit different, but what you see here are all the column names."

Second User Research Session

Following the initial user research session, the app was updated to address the feedback received. Adjustments were made to the timer functionality, screen layout, language translation, and several other features as per the participants' suggestions. A second version of the application was then provided to the participants for the subsequent user research session. In Multimedia Appendix 5, the feedback received during the second user research session is detailed.

In the second session, participants expressed fewer concerns regarding the application, indicating that many of their initial issues had been resolved. However, they still provided valuable feedback on functionality and appearance. One of the main concerns was that the app was not yet fully translated into Dutch; while significant progress had been made, some parts remained in English. Participants emphasized the importance of complete translation before the app could be effectively used by nurses and other care professionals. One participant remarked, "Most of the app has been translated, but some of the texts are still only in English." Another participant added, "This all needs to be translated before we can hand this to nurses."

Participants also suggested that the criteria for showing and hiding fields during observations should be further refined based on relevance, to streamline the data entry process and reduce cognitive load. They discussed various small changes in the appearance of the app to enhance usability and visual appeal. For instance, they recommended that the "star" button used to mark important observations be made a different color to stand out more prominently. One participant suggested, "Could you make the star button orange to make it stand out a bit more?" In addition, participants found the criteria for marking an observation as "complete" to be somewhat confusing. They proposed that once an activity has been set for an observation, it should be automatically marked as complete to provide clearer feedback to the user. One participant explained, "I would prefer it if the observation could be marked as 'complete' when the activity has been set." Participants also recommended offering flexibility in the number of observation rounds that could be added, as different research protocols or care routines might require varying numbers of observations. They suggested making the process more efficient by creating a priority list or extending the default time block for certain observations that typically take longer. Furthermore, participants emphasized the importance of ensuring that the app's manual could be accessed from within the app itself. They stressed that the manual should be clear, comprehensive, and align with users' practical experiences to facilitate ease of use, especially for new users.

By the conclusion of the second user research session, participants expressed optimism about the app's potential to improve their workflow and reduce the time spent on manual data entry and processing. They acknowledged the responsiveness of the development team to their feedback and looked forward to future iterations of the app that would incorporate their latest suggestions.

Discussion

Principal Findings

This study successfully developed an app-based version of the MEDLO-tool aimed at assessing the daily life of residents with

dementia in nursing homes. By using a UCD approach, we incorporated feedback from participants throughout the development process, resulting in a minimum viable product that aligns with the needs and preferences of end-users.

The iterative development process revealed that larger issues, such as bugs and significant feature requests, were identified and addressed in the initial phases. Subsequent user research sessions showed a decrease in the number and severity of issues, indicating progressive refinement of the app. Participants expressed that the app improved upon the Excel-based MEDLO-tool by streamlining data collection and reducing manual input, thereby potentially enhancing usability.

Balancing end-user feedback with established best practices and user interface guidelines was crucial during development [34-37]. While participants provided valuable suggestions, not all could be implemented due to conflicting requests and constraints related to usability standards. For instance, some users desired a more complex timer feature, whereas others preferred simplicity. Decisions were made to benefit the majority and adhere to platform guidelines to ensure familiarity and ease of use [34,35].

The effectiveness of the UCD approach in this study aligns with previous research demonstrating its benefits in software development [38,39]. A key factor in our success was the development team's familiarity with the context of long-term care for older adults. This domain knowledge facilitated insightful conversations between developers and participants, reducing misunderstandings and accelerating the development process. Previous studies have highlighted that cognitive similarity and shared jargon can enhance team performance [40,41].

Limitations

Despite these strengths, the study has limitations that warrant consideration. The participant sample consisted primarily of researchers affiliated with the University, which may limit the generalizability of the findings to care professionals who are the intended end-users in clinical settings. While these researchers had extensive experience with the MEDLO-tool, involving a broader range of stakeholders, including nurses and other care staff, could provide additional insights into usability and practical implementation. Furthermore, the app was not tested in a real-life nursing home environment, which could have revealed context-specific challenges and opportunities.

Future Work

Future research should focus on conducting pilot studies to evaluate the feasibility and effectiveness of the MEDLO app in real-world nursing home settings. Involving care professionals in these studies would help assess the app's usability, accessibility, and impact on daily workflows. In addition, gathering feedback from residents and their families could provide a more comprehensive understanding of the app's influence on care quality.

Further development could explore integrating advanced technologies, such as artificial intelligence and machine learning, to enhance data analysis and provide predictive insights [42,43]. For example, natural language processing could be used to automate the interpretation of observational data, aiding care providers in identifying patterns and potential areas for intervention. However, implementing such technologies would require careful consideration of ethical implications, data privacy, and user acceptance.

Future research may also investigate the usability in practice, by letting nurses use the app in their daily work. This could provide valuable insights into the app's integration into existing workflows and its impact on the quality of care provided to residents. In addition, longitudinal studies could assess the long-term effects of using the app on care outcomes and staff satisfaction. This information could inform further refinements and improvements to the MEDLO app, ensuring its relevance in long-term care.

Conclusion

This study successfully demonstrates the viability and demand for an app-based MEDLO-tool for assessing residents with dementia in nursing homes. By using a UCD approach, this study addressed the limitations of the existing Excel-based system by offering a more efficient and user-friendly alternative. This study shows that involving users early on in the process and keeping them involved can have a positive effect on the usability of an application. The MEDLO app shows that a UCD approach can provide real benefits in the development of a digital tools in nursing homes.

Data Availability

The datasets generated or analyzed during this study are available from the Living Lab in Aging and Long-term Care (ouderenzorg@maastrichtuniversity.nl) on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The interview guide used to the conduct the initial interviews.
[DOCX File, 15 KB - humanfactors_v12i1e57911_app1.docx]

Multimedia Appendix 2

The interview guide used for the user research sessions.
[DOCX File, 14 KB - humanfactors_v12i1e57911_app2.docx]

Multimedia Appendix 3
A screenshot of what the dashboard of the application looks like.
[PNG File, 165 KB - humanfactors_v12i1e57911_app3.png]

Multimedia Appendix 4
Feedback from the initial user research session.
[DOCX File, 16 KB - humanfactors_v12i1e57911_app4.docx]

Multimedia Appendix 5
Feedback from the second user research session.
[DOCX File, 15 KB - humanfactors_v12i1e57911_app5.docx]

References

1. de Boer B, Beerens HC, Zwakhalen SMG, Tan FES, Hamers JPH, Verbeek H. Daily lives of residents with dementia in nursing homes: development of the Maastricht electronic daily life observation tool. Int Psychogeriatr 2016 Aug;28(8):1333-1343. [doi: 10.1017/S1041610216000478]
2. Logsdon RG, Gibbons LE, McCurry SM, Teri L. Assessing quality of life in older adults with cognitive impairment. Psychosom Med 2002 May;64(3):510-519. [doi: 10.1097/00006842-200205000-00016]
3. Dunton GF. Ecological momentary assessment in physical activity research. Exerc Sport Sci Rev 2017 Jan;45(1):48-54. [doi: 10.1249/JES.0000000000000092] [Medline: 27741022]
4. Beerens HC, de Boer B, Zwakhalen SMG, et al. The association between aspects of daily life and quality of life of people with dementia living in long-term care facilities: a momentary assessment study. Int Psychogeriatr 2016 Aug;28(8):1323-1331. [doi: 10.1017/S1041610216000466]
5. Bajenaru L, Marinescu IA, Dobre C, Prada GI, Constantinou CS. Towards the development of a personalized healthcare solution for elderly: from user needs to system specifications. Presented at: 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI); Jun 25-27, 2020; Bucharest, Romania. [doi: 10.1109/ECAI50035.2020.9223254]
6. Stewart AL, King AC. Evaluating the efficacy of physical activity for influencing quality-of-life outcomes in older adults. Ann Behav Med 1991 Sep 1;13(3):108-116. [doi: 10.1093/abm/13.3.108]
7. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 1;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415]
8. Moore RC, Depp CA, Wetherell JL, Lenze EJ. Ecological momentary assessment versus standard assessment instruments for measuring mindfulness, depressed mood, and anxiety among older adults. J Psychiatr Res 2016 Apr;75:116-123. [doi: 10.1016/j.jpsychires.2016.01.011] [Medline: 26851494]
9. Or CKL, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009;16(4):550-560. [doi: 10.1197/jamia.M2888] [Medline: 19390112]
10. Gagnon MP, Desmartis M, Labrecque M, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277. [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
11. Beerens HC, Zwakhalen SMG, Verbeek H, et al. The relation between mood, activity, and interaction in long-term dementia care. Aging & Mental Health 2018 Jan 2;22(1):26-32. [doi: 10.1080/13607863.2016.1227766]
12. Lee H, Chapiro J, Schernthaner R, et al. How I do it: a practical database management system to assist clinical research teams with data collection, organization, and reporting. Acad Radiol 2015 Apr;22(4):527-533. [doi: 10.1016/j.acra.2014.12.002] [Medline: 25641319]
13. Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review. J Med Syst 2016 Dec;40(12):252. [doi: 10.1007/s10916-016-0628-9] [Medline: 27714560]
14. Adipat B, Zhang D. Interface design for mobile applications. Presented at: 11th Americas Conference on Information Systems, AMCIS 2005; Aug 11-14, 2005; Omaha, NE, United States.
15. Martínez-Pérez B, de la Torre-Díez I, López-Coronado M. Mobile health applications for the most prevalent conditions by the World Health Organization: review and analysis. J Med Internet Res 2013 Jun 14;15(6):e120. [doi: 10.2196/jmir.2600] [Medline: 23770578]
16. dos Santos Picanco LC, Ozela PF, de Fatima de Brito Brito M, et al. Alzheimer's disease: a review from the pathophysiology to diagnosis, new perspectives for pharmacological treatment. CMC 2018 Sep 4;25(26):3141-3159. [doi: 10.2174/0929867323666161213101126]

17. Bernstein Sideman A, Harrison KL, Garrett SB, Naasan G, Group D, Ritchie CS. Practices, challenges, and opportunities when addressing the palliative care needs of people living with dementia: Specialty memory care provider perspectives. Alzheimers Dement (N Y) 2021;7(1):e12144. [doi: 10.1002/trc2.12144] [Medline: 33969177]
18. Wang W, Grundy J, Khalajzadeh H, Madugalla A, Obie HO. Designing adaptive user interfaces for mhealth applications targeting chronic disease: a user-centric approach. arXiv. Preprint posted online on May 14, 2024. [doi: 10.48550/arXiv.2405.08302]
19. Norman DA. User Centered System Design: New Perspectives on Human-Computer Interaction: CRC Press; 1986.
20. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985 Mar;28(3):300-311. [doi: 10.1145/3166.3170]
21. Tan L, Szebeko D. Co-designing for dementia: the Alzheimer 100 project. AMJ 2008 Nov 1:185-198. [doi: 10.4066/AMJ.2009.97]
22. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020 Feb 11;18(1):17. [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
23. Martı´nez-Pérez B, Torre-Dı´ez I, López-Coronado M. Privacy and security in mobile health apps: a review and recommendations. J Med Syst Springer 2015 Jan;39(1). [doi: 10.1007/s10916-014-0181-3]
24. McCusker D, Guzik K. User-Centered Design for Middleware: Elsevier; 2007.
25. Couture B, Lilley E, Chang F, et al. Applying user-centered design methods to the development of an mHealth application for use in the hospital setting by patients and care partners. Appl Clin Inform 2018 Apr;9(2):302-312. [doi: 10.1055/s-0038-1645888] [Medline: 29742756]
26. Abras C, Maloney-Krichmar D, Preece J. Bainbridge, W Encyclopedia of Human-Computer Interaction Thousand: Sage Publications; 2004, Vol. 37:445-456.
27. Wilson C. Interview Techniques for UX Practitioners: A User-Centered Design Method: Elsevier; 2013.
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
29. Maguire M, Delahunt B. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. AISHE-J 2017;9(3). [doi: 10.62707/aishej.v9i3.335]
30. Strauss A, Corbin J. Basics of Qualitative Research Techniques: Sage Publications; 1998.
31. Fox S, Brown LJE, Antrobus S, et al. Co-design of a smartphone app for people living with dementia by applying agile, iterative co-design principles: development and usability study. JMIR Mhealth Uhealth 2022 Jan 14;10(1):e24483. [doi: 10.2196/24483] [Medline: 35029539]
32. Your research: is it subject to the WMO or not? Central Committee on Research Involving Human Subjects. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/your-research-is-it-subject-to-the-wmo-or-not [accessed 2025-02-13]
33. Research code version 2024. Maastricht UMC+. 2024. URL: https://www.mumc.nl/sites/default/files/2024-02/Research_Code_v2024.pdf [accessed 2025-02-13]
34. Visnapuu K. Consistency, efficiency, and scalability in design systems: a comparative analysis of material design, Apple's human interface guidelines, and IBM design systems. : Uppsala University; 2023.
35. Gülenman T. Designing better mobile apps: an experimental evaluation of apple's and google's design guidelines: how analysing the human interface guidelines for ios and material design for android better our understanding of the usability challenges app users face and what we can do to overcome key issues [Thesis]. : Jönköping University School of Engineering; 2022 URL: https://www.diva-portal.org/smash/get/diva2:1677778/FULLTEXT01.pdf [accessed 2025-01-31]
36. Eberts RE. User Interface Design: Prentice-Hall, Inc; 1994.
37. Peter Dey P, Raj Sinha B, Amin M, Badkoobehi H. Best practices for improving user interface design. IJSEA 2019 Sep 30;10(5):71-83 [FREE Full text] [doi: 10.5121/ijsea.2019.10505]
38. Mao JY, Vredenburg K, Smith PW, Carey T. The state of user-centered design practice. Commun ACM 2005 Mar;48(3):105-109. [doi: 10.1145/1047671.1047677]
39. Van Elzakker C, Delikostidis I. User-centered design of mobile geo-applications. In: Handbook of Research on Mobile Software Engineering: Design, Implementation, and Emergent Applications: IGI Global; 2013:772-787. [doi: 10.4018/978-1-4666-4301-7.ch037]
40. Kang HR, Yang HD, Rowley C. Factors in team effectiveness: cognitive and demographic similarities of software development team members. Hum Rel 2006 Dec;59(12):1681-1710. [doi: 10.1177/0018726706072891]
41. He J, Butler BS, King WR. Team cognition: development and evolution in software project teams. J Manag Inf Syst 2007 Oct;24(2):261-292. [doi: 10.2753/MIS0742-1222240210]
42. Ashraf A, Imam A. ChatGPT's use case for software engineers. IET Conf Proc 2023 Jul 26;2023(5):487-492. [doi: 10.1049/icp.2023.1537]
43. Roziere B, Gehring J, Gloeckle F, Sootla S, Gat I, Tan XE, et al. Code Llama: open foundation models for code. arXiv. Preprint posted online on Aug 24, 2023. [doi: 10.48550/arXiv.2308.12950]

Abbreviations

EMA: ecological momentary assessment
MEDLO: Maastricht Electronic Daily Life Observation
MEDLO-tool: Maastricht Electronic Daily Life Observation tool
MVP: minimum viable product
UCD: user-centered design
WMO: Medical Research Involving Human Subjects Act

Edited by A Kushniruk, Y Quintana; submitted 29.02.24; peer-reviewed by H Wörtche, ZVD Heijden; revised version received 22.11.24; accepted 25.11.24; published 27.02.25.

Please cite as:
Hacking C, de Boer B, Verbeek H, Hamers J, Aarts S
Developing an App for Real-Time Daily Life Observations in a Nursing Home Setting: Qualitative User-Centered Co-Design Approach
JMIR Hum Factors 2025;12:e57911
URL: https://humanfactors.jmir.org/2025/1/e57911
doi:10.2196/57911

Oncology Clinicians' Perspectives of a Remote Patient Monitoring Program: Multi-Modal Case Study Approach

Ann Marie Mazzella-Ebstein[1], RN, PhD; Robert Daly[2], MBA, MD; Jennie Huang[3], BA; Camila Bernal[4], MPH; Clare Wilhelm[5], PhD; Katherine S Panageas[6], Dr PH; Jessie Holland[7], RN, MSN, NE-BC; Rori Salvaggio[7], RN, MS, NEA-BC; Jill Ackerman[4], MBA; Jennifer Cracchiolo[5], MD; Gilad Kuperman[8], MD, PhD; Jun Mao[9], MSCE, MD; Aaron Begue[10], RN, MS, CNP; Margaret Barton-Burke[1], RN, PhD

[1]Nursing Research, Department of Nursing, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[10]Hospital Administration, Advanced Practice Providers, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[2]Thoracic Services, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[3]Memorial Sloan Kettering Cancer Center, New York, NY, United States

[4]Hospital Administration, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[5]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[6]Epidemiology - Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[7]Nursing Ambulatory Services, Department of Nursing, Memorial Sloan Kettering Cancer Center, New York, NY, United States

[8]Digital Informatics & Technology Solutions (DigITs), Memorial Sloan Kettering Cancer Center, New York, NY, United States

[9]Integrative Medicine Services, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Corresponding Author:
Ann Marie Mazzella-Ebstein, RN, PhD
Nursing Research, Department of Nursing, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Abstract

Background: Remote patient monitoring (RPM) aims to improve patient access to care and communication with clinical providers. Overall, understanding the usability of RPM applications and their influence on clinical care workflows is limited from the perspectives of clinician end users at a cancer center in the Northeastern United States.

Objective: This study aims to explore the usability and functionality of RPM and elicit the perceptions and experiences of oncology clinicians using RPM for oncology patients after hospital discharge.

Methods: The sample included 30 of 98 clinicians (31% response rate) managing at least 5 patients in the RPM program and responding to the mHealth usability between March 2021 and October 2021. Overall, clinicians responded positively to the survey. Item responses with the highest proportion of disagreement were explored further. A nested sample of 5 clinicians who responded to the study survey (30% response rate) participated in interview sessions conducted from November 2021 to February 2022, averaging 60 minutes each.

Results: Survey responses highlighted that RPM was easy to use and learn and verified symptom alerts during follow-up phone calls. Areas to improve identified practice changes from reporting RPM alerts through digital portals and its influence on clinicians' workload burden. Interview sessions revealed 3 main themes: clinician understanding and usability constraints, patient constraints, and suggestions for improving the program. Subthemes for each theme were explored, characterizing technical and functional limitations that could be addressed to enhance efficiency, workflow, and user experience.

Conclusions: Clinicians support the value of RPM for improving symptom management and engaging with providers. Improvements to address RPM challenges include functional changes to enhance the program's utility, such as input from patients about temporal changes in their symptoms and technical resources for home monitoring devices.

(JMIR Hum Factors 2025;12:e60585) doi:10.2196/60585

KEYWORDS

cancer; oncology; clinician end users; remote patient monitoring; digital health; implementation science; patient monitoring; patient access; care; communication; usability; functionality; survey; interview; efficiency; workflow; user; clinician support

Introduction

Improving the patient's access to care and enhancing their quality of life by preventing readmission is the primary goal of posthospital care delivery, yet traditional oncology care models often lack communication and monitoring after discharge [1-5]. Consequently, re-admissions for oncology patients within 30 days of discharge for symptoms that could be mitigated with proactive, remote patient monitoring (RPM) is an opportunity for advancing oncology patient care [6-8]. Our prior research at Memorial Sloan Kettering Cancer Center (MSK) highlights the potential of RPM to facilitate transitions of care through optimized patient-provider communication and proactive engagement through digital technology [7,8]. The RPM program was designed to reduce unplanned care visits following discharge from the hospital by proactively monitoring patient status via daily questionnaires. Electronic patient-reported outcome (ePRO) assessments generated alerts for patients enrolled in the RPM that were sent to the patient's primary oncology care clinician team. Details regarding the early development and integration of the organizational RPM project, as well as the patients recruited have been previously published [3,9-11]. However, limited research exists describing the perspectives and experiences of the clinician end users regarding how digital interventions are integrated into clinical care and the influence of technological health care models on patient care workflows.

To address this gap, our team developed this study to gain a greater understanding of the influence of RPM on clinical workflows in daily practice as part of a larger institutional initiative supporting center-wide symptom management during the transition of care from the inpatient to the outpatient setting [3,9-11]. The purpose of this study is to explore the usability of the RPM program and the perceptions and experiences of oncology clinicians using RPM to care for patients in practice after hospital discharge. The broader aim was to identify the impact, perceived usefulness ease of use, user control of the RPM, and barriers and facilitators experienced during the initial RPM implementation. The Consolidated Framework for Implementation Research [12] guided this study to query oncology clinicians, the key contributors to this management strategy, about their experiences, satisfaction, usefulness, and value of RPM.

Methods

Overview

The study's design used a multi-modal case study approach, using quantitative (survey) and qualitative (interviews and focus groups) to characterize the phenomenon and context of integrating RPM into the practice setting [13-18]. This approach encourages multiple sources of evidence to understand the clinician's perspectives about the use and function of RPM in real-world clinical practice and its influence on patient care workflows.

Initially, survey methods assessed clinicians' responses to the overall usability, functionality, and value of RPM. Second, clinicians participated in semistructured interviews and focus group sessions to elicit feedback and explanations about items of concern found in the survey responses [13,14,17]. The interview sessions aimed to delve deeper into the survey concerns and examine the use of the RPM application from a clinical perspective.

Setting and Clinician Sample

This study focused on clinician end users, specifically nurses involved in clinical office practices that integrated the initial organization's RPM initiative [3,9-11] from October 15, 2018, to July 10, 2019, at MSK in the Northeastern United States. The clinicians in this study were the first group of end users involved in caring for oncology patients using RPM. The RPM program was rolled out sequentially, beginning with the oncology services that care for patients with the highest symptom burden (Table 1). Clinicians from physician office practice settings, who had experience in caring for at least 5 patients enrolled in RPM, were invited to participate in the usability survey. This criterion ensured that the clinicians had ample experience with using the RPM program. Additionally, a nested sample of clinicians who completed the survey were recruited to take part in semistructured interviews to provide further information about the challenges of the RPM identified from the survey [19-22].

Table . Description of practice clinicians by percent of the sample. Clinicians: registered nurse in practice settings; responses (n=30; response rate 31%).

Oncology specialty disease management service	m-Adapted health questionnaire disseminated, n (%)[a]
Thoracic (THR)	23 (24)
Gastro-intestinal (GI)	30 (31)
Breast (BR)	15 (15)
Sarcoma (SAR)	6 (6)
Genito-urinary (GU)	6 (6)
General medicine oncology (GMO)	5 (5)
Head & neck (HN)	4 (4)
Melanoma (MEL)	2 (2)
Myeloma (MYL)	2 (2)
Bone marrow transplant (BMT)	1 (1)
Lymphoma (LYMP)	1 (1)
Leukemia (LEU)	1 (1)
Covering clinicians/no specific service	2 (2)
Total surveys sent	98 (100)

[a]Percentage of the sample.

Data Collection

Survey

Eligible clinicians in this study who met the study criteria were identified using the RPMs dashboard. Clinicians were recruited using the organization's email with a link to a web-based consent and m-Adapted Health Usability Questionnaire from March 2021 to October 2021. To maintain the anonymity of the clinicians, broad demographic data were collected to quantify clinicians' years of experience in their current organizational roles and use RPMs within electronic health record systems.

The m-Adapted Health Usability Questionnaire [23,24], is a 29-item self-report survey, that assessed an overall understanding of the functionality of the RPM from the perspectives of the clinician end users including the quality of work life, perceived usefulness, and ease of use, and user control of RPM in a 7-point Likert Scale (1 [strongly disagree] to 7 [strongly agree]). The questionnaire takes about 13 minutes to complete in REDCap (research electronic data capture; Vanderbilt University) [25]. The Cronbach α for all scale values was >0.7 with scores ranging from 0.85 to 0.92. Permission to use the questionnaire was not required [24].

Interview Sessions and Guide

Clinicians who completed the study's survey were also recruited through the organization's email to take part in 60-minute interviews or focus group sessions. The interview guide was developed by the research team based on understanding areas where the usability of RPM was challenging for the clinicians. Items from the usability survey where the percentage of responders who disagree, disagree, and strongly disagree exceeded 44%, informed the interview and focus group guide. Topics in the guide were broad enough to elicit narrative data describing the background and contextual characteristics related to the clinician's experience with RPM [26,27]. The interview guide was organized into four topics: (1) General understanding and information about the RPM; (2) Experiences, barriers, and challenges with using the RPM; (3) The Influence of the RPM on the current workflow; and (4) Suggestions for improvement (Table 2). Interviews and focus groups were conducted via the web by the PI (AMME) and the qualitative methods specialist (MBB) from November 2021 to February 2022. Verbal consent was obtained from clinicians to record sessions.

Table . Interview guide.

Topic	Main statements	Probe questions	Aligns survey
Topic 1: provide general information about the topic of study and ask questions related to the specific service and the specific patient needs.	• Where do you work? • Can you tell me about your work with RPM[a]? • What can you tell us/each other about RPM?	• Can you briefly describe the patient group in your office practices (patient symptom burden, any factors in this group that need to be addressed through RPM)? • Can you describe your workflow and the technology systems used in your daily work activities? • Can you describe the workflow, and the systems used when providing care? • How does the RPM fit into the care workflow?	Q[b] 3,4
Topic 2: issues and barriers that are specific to using the RPM.	• What are the strengths/barriers to using RPM from the nursing perspective? • What are the strengths/barriers to using RPM from the patient's perspective? • Do you think RPM is valuable: to nurses the patients?	• What functionality is found in the current systems and applications used during patient care activities to make caring for patients easy? • What functionality in the RPM is not working? Describe? • Do you think the patients like using the RPM? • Does the RPM provide value for the patients? Care? • Do you think patient education about the RPM helps?	Q 7,9,10,13
Topic 3: practice workflow; getting to the issue about the RPM, frequency of using the toolkit, and the influence of workflow.	• How does the RPM toolkit work with your workflow? • How do you use the RPM apps to provide care? • How do you use the RPM toolkit using the digital apps?	• Describe how you use the RPM (and patients with remote monitoring devised, ie, pulse oximetry) • How do you use the information in caring for your patients? • Can you describe the communication among practitioners using the RPM?	Q 18 - 20
Topic 4: functionality, if modified, would be meaningful and helpful for the RPM, improved workflow, user satisfaction, and improved patient outcomes.	• What are your thoughts about how RPM functions? • How could it be improved?	• If you could fix anything about the RPM, what would it be? • To what extent do you think the modifications described could influence patient outcomes that patient care? • What current functionality has the greatest influence on workflow?	Q 23 - 25

[a]RPM: remote patient monitoring.

[b]Q: question.

Data Analysis

Survey

Survey response data were extracted from REDCap [25] to a deidentified Excel spreadsheet before analysis. To ensure anonymity, demographic information was only provided for clinicians who were sent the surveys and not collected for clinicians who responded. Data responses were initially scored in REDCap [25] for the frequency and proportion of agreement or disagreement with survey items consistent with the 7-point response scale. The scale responses were then collapsed into 3 groups, disagreement (strongly disagree, disagree), neutral (neither agree nor disagree), and agreement (agree and strongly agree). The Strengthening the Reporting of Observational Studies in Epidemiology guidelines were used to report the study's quantitative findings [28].

Interviews Sessions

Data from the interviews and focus groups followed the Consolidated Criteria for Reporting Qualitative Research

guidelines and were used to report the study's findings [29]. The rigor and validity of the data were supported by clarifying participants' statements during interview sessions. The transcripts were audio recorded, transcribed verbatim, and independently reviewed by 2 reviewers using thematic content analysis [30]. Themes and subthemes were constructed based on the verbal responses from the clinicians. Team consensus was determined by agreeing upon the best representation of the data [24,26,27,30].

Ethical Considerations

This research is part of an ongoing organizational program of RPM initiatives. This research obtained ethical approval from the MSK Institutional Review Board (X20-086) as exempt research and follows the ethical principles and guidelines of the Belmont Report. All responders to surveys completed web-based informed consent to participate. Participants in focus groups and interviews provided informed verbal consent.

Results

Overview

A total of 35 clinicians participated in this study from both survey and interview sessions. Of the 98 clinicians who had experiences with at least 5 patients enrolled in RPM and received the study's survey, 30 responded (31% response rate). A nested sample of 5 clinicians (17% response rate) from this original group of 30 clinicians who completed the survey, also agreed to take part in 4 sessions (3 interviews; 1 focus group of 2 clinicians). Demographic data were not collected with the survey to maintain the anonymity of the clinician responders. However, the majority of clinicians who responded to the survey and interview sessions represented the Gastrointestinal, Thoracic, and Breast oncology services known to have the highest symptom burden. Of the clinicians who took part in the interview sessions, 3 clinicians had 5 or more years of organizational work experience, while 2 clinicians had 2 years or less of experience.

Survey Responses

Overall, the clinicians reported that the RPM was easy to use and learn and that symptoms communicated through the RPM program were confirmed during follow-up phone calls. Suggested areas for improvement included communication and practice changes related to symptom management and the clinicians' workload burden. Findings highlighting survey items with the highest frequency and proportion of survey responses are presented (Table 3).

Table . Frequency distribution of m-Adapted health questionnaire responses (n=30)[a].

Statements about remote patient monitoring (RPM)	Total responses, n	Agree/somewhat agree/strongly agree, n (%)[b]	Neither agree nor disagree, n (%)	Disagree/somewhat disagree/strongly disagree, n (%)	Unknown, n
Ease of use and functionality as easy to learn					
1. RPM is easy to use	27	18 (67)	2 (7)	7 (26)	3
2. RPM is easy for me to learn	27	18 (67)	7 (15)	5 (18)	3
3. Liked the digital interface of the patient data received through portal secure message alerts	26	8 (31)	4 (15)	14 (54)	4
4. I liked the digital interface of the patient data received through the Splunk/summary dashboard	26	7 (27)	12 (46)	7 (27)	4
5. Information in the Splunk/summary dashboard was well-organized	25	12 (48)	9 (36)	4 (16)	5
Integrating the RPM into workflows					
6. RPM has usable functions and capabilities	27	14 (52)	6 (22)	7 (26)	3
7. RPM has been appropriate for me to care for patients	27	11 (44)	1 (4)	14 (52)	3
8. Easy to integrate into my current clinical workflow	27	11 (41)	4 (15)	12 (44)	3
Acceptable for practice					
9. An acceptable way to coordinate health care services	26	9 (35)	5 (19)	12 (46)	4
10. Improved communication between my colleague's office practice teams for patient symptoms	26	10 (38)	1 (4)	15 (58)	4
11. Prompts me to refer patients to a specialist for symptom management	25	9 (36)	5 (20)	11 (44)	5
12. RPM is useful for my health care practice	26	12 (46)	4 (15)	10 (39)	4
13. Improved my ability to deliver health care services	27	7 (26)	4 (15)	16 (59)	3
14. Helped me manage my patient's symptoms effectively	27	11 (41)	5 (18)	11 (41)	3

Statements about remote patient monitoring (RPM)		Total responses, n	Agree/somewhat agree/strongly agree, n (%)[b]	Neither agree nor disagree, n (%)	Disagree/somewhat disagree/strongly disagree, n (%)	Unknown, n
RPM convenience						
	15. Is convenient for me to communicate with patients	26	10 (39)	4 (15)	12 (46)	4
	16. Had many more opportunities to interact with patients	27	11 (41)	4 (15)	12 (44)	3
	17. Felt comfortable communicating with my patients about symptoms using portal secure messaging	26	12 (46)	1 (4)	13 (50)	4
Devices and symptom management						
	18. Highlighted the high-risk symptoms (not pulse oximetry) provided the correct corresponding severity level for the patient-reported symptoms	27	10 (37)	5 (19)	12 (44)	3
	19. Highlighted symptoms related to pulse oximetry appropriately	27	10 (37)	9 (33)	8 (30)	3
	20. Patients were appropriately identified for pulse oximeter monitoring.	27	12 (44)	7 (26)	8 (30)	3
	21. Pulse oximeter monitoring enabled me to more effectively manage my patient's symptoms	26	10 (38)	8 (31)	8 (31)	4
	22. High-risk symptom alerts were confirmed upon communication with the patient by telephone	27	20 (74)	2 (7)	5 (19)	3
Satisfaction, value, and recommendations						
	23. Using RPM has improved my job satisfaction	27	3 (11)	9 (33)	15 (56)	3
	24. Using RPM decreased my workload	26	2 (11)	2 (8)	21 (81)	4
	25. Adds value to how I can care for my patients	27	8 (30)	7 (26)	12 (44)	3
	26. Patients reported the value of participating in RPM	27	8 (30)	7 (26)	12 (44)	3

Statements about remote patient monitoring (RPM)		Total responses, n	Agree/somewhat agree/strongly agree, n (%)[b]	Neither agree nor disagree, n (%)	Disagree/somewhat disagree/strongly disagree, n (%)	Unknown, n
	27. Overall, I am satisfied with the using RPM	27	9 (33)	4 (15)	14 (52)	3
	28. I would use RPM again to monitor the symptoms of patients	27	10 (37)	6 (22)	11 (41)	3
	29. I would recommend the RPM to my colleagues	26	8 (31)	5 (19)	13 (50)	4

[a]Number of participants who answered the survey.

[b]Percentage of the group responses.

Usability

The RPM was easy for clinicians to use but was influenced by a shift in basic assumptions with patients now instructed to report symptoms through portal messages rather than calling the medical offices (14/26, 54%).

Integrated Workflows

The proportion of responses was fairly divided between agreement and disagreement. While 14 (52%) clinicians agreed that the RPM had usable functions and capabilities, 14 (52%) clinicians disagreed that the RPM was suitable for their patient care needs.

Acceptability in Practice

A total of 16 (59%) clinicians reported a disagreement with the notion that RPM improved their ability to care for their patients and, 12 (46%) clinicians reported that the RPM was useful in their practice.

Convenience in Patient Care

Despite similar proportions of agreement and disagreement overall, a higher proportion of clinicians (13/26, 50%) disagreed with feeling comfortable communicating with patients about symptoms through portal messages.

Symptom Management

Clinicians reported that devices were appropriately used for 12 (44%) patients and that high-risk symptoms were confirmed during follow-up phone calls to 20 (74%) patients. The responses were fairly split between agreement (10/27, 37%) and disagreement (12/27, 44%) on how well the devices provided severity levels. Additionally, 10 (38%) patients agreed that RPM helped them manage their patients.

Satisfaction and Value

A total of 21 (81%) clinicians reported that the RPM did not decrease their workload and 15 (56%) clinicians reported it did not improve job satisfaction. In addition, 14 (52%) of the clinicians were not satisfied with RPM, and 13 (50%) clinicians would not recommend the RPM to colleagues. The proportion of clinicians who neither agreed nor disagreed was similar to those who reported disagreement and was further explored in the interview sessions.

Interview Sessions

The interviews and focus group sessions consistently provided similar information about the RPM thus achieving thematic saturation [30]. All clinicians reported a limited understanding of the RPM during its implementation and suggested that ongoing educational modules and supporting technical support would enhance the RPM program for the clinician end users. Three major themes emerged from the interviews: clinician understanding and usability constraints, patient constraints, and suggestions for improvement. Subthemes were further explored (Table 4).

Table . Themes and subthemes.

Theme	Subthemes
Theme 1: Clinician understanding and usability constraints	• Clinician uncertainty • Repetitive alerts • Alignment with clinician workflows • Program value
Theme 2: Patient constraints	• Appropriate patient enrollment • Timing of enrollment • Communication during and after enrollment
Theme 3: Suggestions for improvement	• Clinical champions needed • Program modifications • Information technology support for end users

Theme 1: Clinician Understanding and Usability Constraints

Overview

Theme 1 encompasses the information given to the clinicians during their RPM orientation and how they applied the application to their current workflow. Three subthemes included clinician uncertainty, repetitive alerts, alignment with clinician workflows, and program value.

Subtheme: Clinician Uncertainty

The RPM was rolled out in stages starting with oncology services known to have patients with high symptom burdens. The clinicians involved in the initial rollout reported more knowledge and understanding about the RPM compared with clinicians who were involved later in the rollout. Clinicians joining later in the rollout reported having little training about the RPM. However, they reported that the increased frequency of using the RPM helped them navigate the program for addressing symptom management alerts and functionality.

> *I have only worked with it (RPM) for a couple of patients in the outpatient setting, but I get a notification (through) the portal … to notify us that the patient has enrolled in the RPM program…During this timeframe patients (complete) a survey every day about their symptoms and how they feel.* [R32]

Subtheme: Repetitive Alerts

All clinicians reported concerns about repetitive alerts from daily patient surveys. Each day that a patient reported a symptom and generated an alert [1,3,10], clinicians called the patient to verify it and confirm if the symptom was worse, better, or the same as the previous survey. Although they could see past patient response trends in the system, this required substantial effort and did improve the patient's care for symptoms. When alerts were consistently reported, the clinicians used their judgment to decide when to call patients for verification.

> *There are a couple of categories that we feel were a little monotonous. ie, was (the pain/symptom) worse? If they are at stage four lung cancer, (the patient) is on treatment and they have fatigue, every day… (the patient cannot modify) their answers, (for example) moderate fatigue; I am having trouble doing my activities of daily living, so they are clicking that every day.* [R11]

Subtheme: Aligning With Clinician Workflow

Clinicians reported that before implementing the RPM, patients were instructed to call the physician's medical offices to report symptoms. Although none of the clinicians reported significant changes to their clinical workflows, they all raised concerns about the shift from patients calling the medical oncology office for unrelieved symptoms through a digital portal. This change made clinicians apprehensive about potentially missing symptoms reported in patient portals.

> *Before RPM, patients were not supposed to report symptoms through the portal, but a lot of patients, ended up just doing that because they were home.* [R11]

Subtheme: Program Value

Clinicians reported that the RPM was of immense value to them and their patients and supported its inclusion in the organization's future care delivery. Patients liked that their office practice clinician proactively contacted and interacted with them after they were discharged from the hospital.

> *I do not think they (the patients) mind and love to be followed up closely, I mean my patients would love for me, a call them every day. I think patients prefer (clinicians) calling them over calling the office.* [R41]

Theme 2: Patient Constraints

Overview

The patient constraint themes involved the transition of care from the inpatient setting to RPM following discharge. Three subthemes emerged: the appropriate patient enrollment, the timing of enrollment, and communication during and after enrollment.

Subtheme: Appropriate Patient Enrollment

Nurses from the discharging inpatient unit were responsible for educating patients about using RPM and program enrollment. However, office practice clinicians stated that not all patients enrolled by the discharge team were appropriate due to a lack of technical proficiency, ability, and understanding of the purpose of the RPM. Clinicians consistently reported that they should be included in RPM enrollment decisions for their patients.

> *There is a disconnect, (between enrolling the patient and educating them about the program) and sending symptoms via the portal messaging using the patient portal. The policy is that we are given two business days to answer portal messages…clearly, we cannot do that when it is a separate message. We need to speak to that person (By phone).* [R21]

Subtheme: Timing of Enrollment

Office practice clinicians reported that the time of discharge is overwhelming for patients and inpatient unit teams have limited time to prepare the patients to leave the hospital. During this study, pandemic-related social distancing restrictions prevented caregivers and family members from taking part in the discharge process, further complicated RPM enrollment.

> *The second thing that I wanted to suggest for patients wanting to quickly leave the hospital and be discharged and suddenly just get bombarded with all this information, is to discuss the RPM program the day (before) discharge.* [R32]

Subtheme: Communication During and After Enrollment

Clinicians reported that a patient's age was not a factor in RPM enrollment, as older adults familiar with technology could engage digitally with providers. However, patients with language barriers or limited access to electronic devices needed assistance from caregivers or family members similar to findings in a prior

study [8]. The RPM assessments were intended to be completed during weekday work hours, from 9 AM to 5 PM Monday through Friday. Clinicians described receiving the surveys late in the day, or after hours added stress for the nursing team to address alerts.

> *We have a lot of foreign-speaking patients ---, you know it is not them who are filling out the survey ... and it may be one of their relatives or their daughter. But the daughter may not live with them ... so how do you know what is going on?* [R11]

Theme 3: Suggestions for Improvement

Overview

The Clinicians supported using RPM to sustain patient care posthospital discharge. From their perspective, 3 subthemes emerged, the need for clinical champions, technology support for end users (clinicians and patients), and specific RPM modifications.

Subtheme: Clinical Champions Needed

Clinicians expressed the need for continuing education and updates on RPM improvements, infrequent use required relearning the system for each new patient. While there were few Clinical Champions during the initial rollout, additional experts would have helped them navigate the program and supported novice clinicians. They also suggested developing informative slides or videos addressing specific program issues.

> *I think nursing (clinician end-user) … needs more education on the portal because it (the portal) is such a big part of our job. More training is needed for using (connecting) pulse ox to an iPhone (or other devices ie, Android).* [R11]

Subtheme: Program Modifications

Clinicians reported that patient enrollment before hospitalization or early in their hospital admission would improve the transition from inpatient to outpatient care. They also requested allowing patients to change or clarify their symptom responses within the assessment. Frequent patient alerts were seen as potentially problematic [31] and suggested that enabling patients the option to modify their assessment responses would provide more accurate symptom information to the clinical team.

> *There should be a way for them (patients) to say (respond in the survey) no changes or something… so that we don't get the same exact thing (response) that we spoke to the patient yesterday---about because it's not realistic, that I call the patient every day to talk to them about the same time (symptom).* [R42]

Subtheme: Information Technology Support for End Users

Clinicians raised concerns about the lack of integrated IT systems they use for delivering patient care. Additionally, clinicians expressed the need for greater IT support for both patients and staff. Clinicians conveyed concern about the difficulty in resolving connectivity issues and requested dedicated assistance from IT teams.

> *The patients are calling in a panic and you are trying to walk them through over the phone how to set up the device (RPM on patient's device ie, phone, tablet). If it was not set up correctly on discharge, or the device is just not working. (These issues) add more stress to the patient, but it also adds stress to you (the clinician) This becomes the added work.* [R11]

Discussion

Principal Results

This study presented the perceptions and experiences of oncology clinicians when caring for their patients using RPM after hospital discharge. Initially, clinician survey responses highlighted concerns about their understanding, perceptions, and challenges when using RPM as well as its influence on their clinical workflows. These topics were further explored in interview sessions and revealed 3 themes: clinician understanding and usability constraints, patient constraints, and suggestions for improvement. Most clinicians found the RPM easy to use and learn, allowing them to confirm the patient's reported symptoms during follow-up phone calls. They supported RPM's value for both clinicians and patients as a care delivery method in oncology practice, aligning with findings from other organizational studies using ePROs [11,31-36], as well as the challenges and limitations of addressing symptom alerts after the patient transitions from an inpatient to the home setting [7,35,36].

Challenges and Opportunities From the Clinical Setting

This research further emphasized multiple challenges faced by clinicians during the initial RPM roll-out. A major concern was the practice change in clinical workflows whereby patients reported symptoms through electronic portals rather than contacting the medical offices by phone. Despite this change, clinicians felt their comfort level would improve with more experience, and as RPMs aligned with their workflows [7]. An opportunity to improve clinician confidence includes continuing educational tutorials from the initial RPM orientation and more technical and clinical support for both patients and clinicians. These efforts would facilitate aligning clinician workflows with clinical practice, thus leading to reduced stress and improved job satisfaction when caring for oncology patients using RPM [7].

Another challenge was patient constraints and their lack of understanding of RPM, which clinicians suggested might contribute to their uncertainty in digital symptom management reporting. Some patients completed surveys late in the clinic hours making it difficult to address symptoms. In many instances, clinicians reported that family members completed the surveys instead of the patients which raised concerns about the integrity of the symptom alerts. Further research could enhance the understanding of logistical limitations involving patients and caregivers with completing postdischarge symptom assessments at home. A similar study recommended that concerns revealed by patients and caregivers postdischarge could be included in future clinical outcomes [37]. Additionally, the timeframe for enrolling patients in RPM was another

constraint identified by clinicians. Although older patients were comfortable using the RPM [38,39], enrollment at discharge was considered suboptimal. They suggested educating patients in a relaxed environment or office practices before hospitalizations would benefit patients and improve the enrollment process.

A major opportunity for RPM improvement involves addressing repetitive alerts from the daily ePRO assessments, also identified in other ePRO studies [23]. Clinicians verify all patient-reported symptom alerts. When prior symptoms are stable, these alerts could be averted to reduce their workflow burden [40,41]. To enhance effective communication with the patients, clinicians recommended adding context about symptom changes relative to the prior day's ePRO assessment to improve the integrity of alerts.

The perspectives of all clinicians involved in this study expressed a need for additional resources to improve the effectiveness of RPM and their ability to clinically support it. The RPM was initially implemented using a core team of designated clinicians for addressing patient alerts. Studies of RPM cited the inclusion of core teams for this purpose which may be a preferred approach for clinicians in other studies [42]. However, clinicians in this research preferred a hybrid RPM program with primary care clinicians in the clinical practices addressing alerts on weekdays and centralized after-hours and on weekends. Clinicians proposed that leaders provide consistent status updates and education about RPM functionality through tutorial videos in an accessible location. Additional technological support could improve data collection and relieve clinicians of the burden of providing technical support for patients at home [41,42], which is not the best use of their skills [43].

Strengths and Limitations

The consistent reports from survey responses and interviewed clinicians strengthen the study's findings and provide pragmatic recommendations for workflow redesign and enhancing access to care for oncology patients postdischarge [1,2,7,31], As a result of this study, selected modifications were implemented into practice to improve the program's experience for both clinicians and patients. These results provide important foundational work for future pragmatic trials and implementation science to enhance the usability and value of RPM in oncology practice.

The sample was limited to only clinicians who had experiences with at least 5 patients enrolled in the RPM, but typical for case studies [13-18]. They expressed strong positive or negative experiences with the RPM which may have created a selection bias. The study's design was multi-modal which could have included both interviewer and report biases. Clinicians were also employees of the organization and therefore specific demographics were not collected to ensure their anonymity. This study was conducted at one comprehensive cancer center, limiting the generalizability of results. However, the perspectives of clinician end users from oncology services with known symptom burdens can apply to similar RPM implementation initiatives in other organizations worldwide.

Conclusion

The National Cancer Institute calls for studies about cancer-related interventions, which also include program effectiveness from the perspectives of both patients and clinician end users [44]. This study contributes to the National Cancer Institute's initiatives, demonstrating that RPM is a valuable method for communicating with clinical providers and managing patient symptoms during transitions of care from inpatient to the home setting.

Acknowledgments

The authors want to thank the clinicians who participated in this study for their insight and dedication to oncology practice.

Authors' Contributions

AMME drafted the nursing perspective aspect of the proposal and conducted the interviews, transcribed and analyzed the interview data, and drafted the original manuscript. MBB conducted interviews with AMME, reviewed the transcript data, and edited the manuscript. RD is the principal investigator of the organizational protocol involving remote patient monitoring and reviewed and edited the manuscript. J Huang, CB, and JA are involved in providing protocol support for the daily activities of this study. CW was involved with drafting the manuscript and editing the study proposal. KSP, GK, and JM were involved with the organizational protocol and its development, and manuscript review. J Holland, RS, JC, and AB are involved with the organizational project development and review of the manuscript.

Conflicts of Interest

RD reports grant support from the National Institutes of Health and the Emerson Collective, participation on a data safety monitoring board or advisory board with Varian Medical Systems, and stock or stock options in Roche. The remaining authors do not have conflicts of interest to declare.

References

1. Basch E, Deal AM, Kris MG, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565. [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]

2. Basch E, Deal AM, Dueck AC, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197. [doi: 10.1001/jama.2017.7156]
3. Daly B, Kuperman G, Zervoudakis A, et al. InSight care pilot program: redefining seeing a patient. JCO Oncol Pract 2020 Oct;16(10):e1050-e1059. [doi: 10.1200/OP.20.00214]
4. Denis F, Basch E, Septans AL, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. JAMA 2019 Jan 22;321(3):306-307. [doi: 10.1001/jama.2018.18085] [Medline: 30667494]
5. Mooney KH, Beck SL, Wong B, et al. Automated home monitoring and management of patient-reported symptoms during chemotherapy: results of the symptom care at home RCT. Cancer Med 2017 Mar;6(3):537-546. [doi: 10.1002/cam4.1002] [Medline: 28135050]
6. Innovation: oncology care first model: informal request for information. Centers for Medicare and Medicaid Services. 2021. URL: https://innovation.cms.gov/files/x/ocf-informalrfi.pdf [accessed 2024-12-20]
7. Hassett MJ, Cronin C, Tsou TC, et al. eSyM: an electronic health record-integrated patient-reported outcomes-based cancer symptom management program used by six diverse health systems. JCO Clin Cancer Inform 2022 Jan;6:e2100137. [doi: 10.1200/CCI.21.00137] [Medline: 34985914]
8. Doolin JW, Berry JL, Forbath NS, et al. Implementing electronic patient-reported outcomes for patients with new oral chemotherapy prescriptions at an academic site and a community site. JCO Clin Cancer Inform 2021 Dec;5(5):631-640. [doi: 10.1200/CCI.20.00191]
9. Daly RM, Cracchiolo JR, Huang J, et al. Remote symptom monitoring after hospital discharge. JCO 2022 Jun 1;40(16_suppl):1517-1517. [doi: 10.1200/JCO.2022.40.16_suppl.1517]
10. Simon BA, Assel MJ, Tin AL, et al. Association between electronic patient symptom reporting with alerts and potentially avoidable urgent care visits after ambulatory cancer surgery. JAMA Surg 2021 Aug 1;156(8):740-746. [doi: 10.1001/jamasurg.2021.1798] [Medline: 34076691]
11. Daly B, Cracchiolo J, Holland J, et al. Digitally enabled transitional care management in oncology. JCO Oncol Pract 2024 May;20(5):657-665. [doi: 10.1200/OP.23.00565] [Medline: 38382002]
12. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 7;4(1):50. [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
13. Yin RK. Case study evaluations: a decade of progress? In: Evaluation Models: Viewpoints on Educational and Human Services Evaluation: Dordrecht: Springer Netherlands; 2000:185-193. [doi: 10.1007/0-306-47559-6_11]
14. Yin RK. Case Study Research and Applications: Sage; 2018.
15. Hoon C. Meta-synthesis of qualitative case studies: an approach to theory building. Org Res Methods 2013 Oct;16(4):522-556. [doi: 10.1177/1094428113484969]
16. Hollweck T. Robert K. Yin. (2014). Case Study Research Design and Methods (5th ed.). Can J Program Eval 2016;30(1):282. [doi: 10.3138/cjpe.30.1.108]
17. Moss S. Case studies - the YIN approach. Charles Darwin University. 2020. URL: https://www.cdu.edu.au/files/2020-07/Introduction%20to%20case%20studies%20-%20the%20Yin%20approach.docx [accessed 2024-04-30]
18. Baškarada S. Qualitative case study guidelines. TQR 2014;19(40):1-25. [doi: 10.46743/2160-3715/2014.1008]
19. Sargeant J. Qualitative research part II: participants, analysis, and quality assurance. J Grad Med Educ 2012 Mar;4(1):1-3. [doi: 10.4300/JGME-D-11-00307.1] [Medline: 23451297]
20. Boddy CR. Sample size for qualitative research. QMR 2016 Sep 12;19(4):426-432. [doi: 10.1108/QMR-06-2016-0053]
21. Ayton D. Choose your qualitative study qualitative research–a practical guide for health and social care researchers and practitioners. Open Educational Resources Collective. 2023. URL: https://oercollective.caul.edu.au/qualitative-research/chapter/__unknown__-s2/ [accessed 2024-04-17]
22. Pirani SA. Navigating the complexity of sample size determination for robust and reliable results. IJMRR 2024;3(2):73-86. [doi: 10.56815/IJMRR.V3I2.2024/73-86]
23. Schnall R, Cho H, Liu J. Health information technology usability evaluation scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 5;6(1):e4. [doi: 10.2196/mhealth.8851] [Medline: 29305343]
24. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500. [doi: 10.2196/11500] [Medline: 30973342]
25. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
26. Roller MR. The in-depth interview methods: 12 articles on design and implementation. Roller Research. URL: http://rollerresearch.com/MRR%20WORKING%20PAPERS/IDI%20Text%20April%202020.pdf [accessed 2024-12-20]
27. Creswell JW, Creswell JD. Research Design: Qualitative, Quantitative, and Mixed-Methods Approaches: Sage; 2014.

28. Vandenbroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology (Sunnyvale) 2007 Nov;18(6):805-835. [doi: 10.1097/EDE.0b013e3181577511]
29. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
30. Cernasev A, Axon DR. Research and scholarly methods: thematic analysis. J Am Coll Clin Pharm 2023 Jul;6(7):751-755. [doi: 10.1002/jac5.1817]
31. Basch E, Stover AM, Schrag D, et al. Clinical utility and user perceptions of a digital system for electronic patient-reported symptom monitoring during routine cancer care: findings from the PRO-TECT trial. JCO Clin Cancer Inform 2020 Oct;4(4):947-957. [doi: 10.1200/CCI.20.00081] [Medline: 33112661]
32. Daly B, Lauria TS, Holland JC, et al. Oncology patients' perspectives on remote patient monitoring for COVID-19. JCO Oncol Pract 2021 Sep;17(9):e1278-e1285. [doi: 10.1200/OP.21.00269] [Medline: 34085536]
33. Daly B, Nicholas K, Flynn J, et al. Analysis of a remote monitoring program for symptoms among adults with cancer receiving antineoplastic therapy. JAMA Netw Open 2022 Mar 1;5(3):e221078. [doi: 10.1001/jamanetworkopen.2022.1078] [Medline: 35244701]
34. Sandhu S, King Z, Wong M, et al. Implementation of electronic patient-reported outcomes in routine cancer care at an academic center: identifying opportunities and challenges. JCO Oncol Pract 2020 Nov;16(11):e1255-e1263. [doi: 10.1200/OP.20.00357]
35. Keith RE, Crosson JC, O'Malley AS, Cromp D, Taylor EF. Using the consolidated framework for implementation research (CFIR) to produce actionable findings: a rapid-cycle evaluation approach to improving implementation. Implement Sci 2017 Feb 10;12(1):15. [doi: 10.1186/s13012-017-0550-7] [Medline: 28187747]
36. Lawrence K, Singh N, Jonassen Z, et al. Operational implementation of remote patient monitoring within a large ambulatory health system: multimethod qualitative case study. JMIR Hum Factors 2023;10:e45166. [doi: 10.2196/45166PMID]
37. Dowzicky PM, Shah AA, Barg FK, Eriksen WT, McHugh MD, Kelz RR. An assessment of patient, caregiver, and clinician perspectives on the post-discharge phase of care. Ann Surg 2021;273(4):719-724. [doi: 10.1097/SLA.0000000000003479]
38. Melstrom LG, Zhou X, Kaiser A, et al. Feasibility of perioperative remote monitoring of patient-generated health data in complex surgical oncology. J Surg Oncol 2023 Jan;127(1):192-202. [doi: 10.1002/jso.27106] [Medline: 36169200]
39. Beaney T, Clarke J. Home oxygen monitoring and therapy: learning from the pandemic. Curr Opin Crit Care 2023 Feb 1;29(1):34-39. [doi: 10.1097/MCC.0000000000001010] [Medline: 36539965]
40. Handley NR, Schuchter LM, Bekelman JE. Best practices for reducing unplanned acute care for patients with cancer. J Oncol Pract 2018 May;14(5):306-313. [doi: 10.1200/JOP.17.00081] [Medline: 29664697]
41. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251. [doi: 10.1016/S1470-2045(20)30021-8] [Medline: 32359500]
42. Coffey JD, Christopherson LA, Glasgow AE, et al. Implementation of a multisite, interdisciplinary remote patient monitoring program for ambulatory management of patients with COVID-19. NPJ Digit Med 2021 Aug 13;4(1):123. [doi: 10.1038/s41746-021-00490-9] [Medline: 34389787]
43. Offodile AC II, Seitz AJ, Peterson SK. Digital health navigation: an enabling infrastructure for optimizing and integrating virtual care into oncology practice. JCO Clin Cancer Inform 2021 Dec;5(5):1151-1154. [doi: 10.1200/CCI.21.00111]
44. Division of cancer control and population sciences. National Cancer Institute. URL: https://cancercontrol.cancer.gov/ [accessed 2024-05-22]

Abbreviations

ePRO: electronic patient-reported outcome
MSK: Memorial Sloan Kettering Cancer Center
REDCap: research electronic data capture
RPM: remote patient monitoring

Edited by A Kushniruk, E Borycki; submitted 03.06.24; peer-reviewed by J Sharp, L Zheng; accepted 23.09.24; published 24.01.25.

Please cite as:
Mazzella-Ebstein AM, Daly R, Huang J, Bernal C, Wilhelm C, Panageas KS, Holland J, Salvaggio R, Ackerman J, Cracchiolo J, Kuperman G, Mao J, Begue A, Barton-Burke M
Oncology Clinicians' Perspectives of a Remote Patient Monitoring Program: Multi-Modal Case Study Approach
JMIR Hum Factors 2025;12:e60585
URL: https://humanfactors.jmir.org/2025/1/e60585
doi:10.2196/60585

Evaluating the Safety and Usability of an Over-the-Counter Medical Device for Adults With Mild to Moderate Hearing Loss: Formative and Summative Usability Testing

Megan Elizabeth Salwei[1], PhD; Shilo Anders[1], PhD; Carrie Reale[1], MSN, RN, NI-BC; Jason M Slagle[1], PhD; Todd Ricketts[2], PhD; Matthew B Weinger[1], MD, MS

[1]Center for Research and Innovation in Systems Safety, Department of Anesthesiology, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN, United States

[2]Department of Hearing and Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
Megan Elizabeth Salwei, PhD
Center for Research and Innovation in Systems Safety, Department of Anesthesiology, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 800, Nashville, TN, United States

Abstract

Background: Only 15% of the nearly 30 million Americans with hearing loss use hearing aids, partly due to high cost, stigma, and limited access to professional hearing care. Hearing impairment in adults can lead to social isolation and depression and is associated with an increased risk of falls. Given the persistent barriers to hearing aid use, the Food and Drug Administration issued a final rule to allow over-the-counter hearing aids to be sold directly to adult consumers with perceived mild to moderate hearing loss at pharmacies, stores, and online retailers without seeing a physician or licensed hearing health care professional.

Objective: We evaluated the safety and usability of an over-the-counter hearing aid prior to Food and Drug Administration approval and market release.

Methods: We first conducted a formative usability test of the device and associated app with 5 intended users to identify outstanding safety and usability issues (testing round 1). Following design modifications, we performed a summative usability test with 15 intended users of the device (testing round 2). We concurrently conducted a test with 21 nonintended users (ie, users with contraindications to use) to ascertain if consumers could determine when they should not use the device, based on the packaging, instructions, and labeling (testing round 3). Participants were asked to complete 2 - 5 tasks, as if they were using the hearing aid in real life. After each task, participants rated the task difficulty. At the end of each session, participants completed a 10-question knowledge assessment and the System Usability Scale and then participated in debriefing interviews to gather qualitative feedback. All sessions were video recorded and analyzed to identify use errors and design improvement opportunities.

Results: Usability issues were identified in all 3 usability testing rounds. There were minimal safety-related issues with the device. Round 1 testing led to several design modifications which then increased task success in round 2 testing. Participants had the most difficulty with the task of pairing the hearing aids to the cell phone. Participants also had difficulty distinguishing the right and left earbuds. Nonintended users did not always understand device contraindications (eg, tinnitus and severe hearing loss). Overall, test findings informed 9 actionable design modifications (eg, clarifying pairing steps and increasing font size) that improved device usability and safety.

Conclusions: This study evaluated the usability and safety of an over-the-counter hearing aid for adults with mild to moderate hearing loss. Human factors engineering methods identified opportunities to improve the safety and usability of this direct-to-consumer medical device for individuals with perceived mild-moderate hearing loss.

(JMIR Hum Factors 2025;12:e65142) doi:10.2196/65142

KEYWORDS

usability; human factors; patient safety; over-the-counter hearing aids; direct-to-consumer hearing aids; medical device; hearing loss; adult; hearing impairment; hearing aid use; hearing care; formative usability test; safety; mobile phone

Introduction

Approximately 1 in 4 people older than 12 years of age in the United States have hearing loss in 1 or both ears [1], and 30 million American adults could benefit from hearing aid use [2]. Hearing loss is associated with an increased risk for falls [3] and may lead to cognitive decline [4] and dementia [5]. Yet, only 1 in 3 older adults [6] and 15% of all adults with hearing loss use hearing aids [7]. Some barriers preventing adoption

include lack of access, stigma, and financial challenges with traditional hearing aids costing between US $1000 and US $6000 per ear [8]. Aiming to mitigate these barriers, the Food and Drug Administration (FDA) issued a final rule in 2022 [9] permitting the sale of hearing aid devices without a prescription or medical exam. Yet, there are few studies of the usability and safety design challenges of over-the-counter medical devices intended to be sold widely in consumer markets.

Over-the-counter or direct-to-consumer medical devices do not require a prescription and can be sold at retail outlets including pharmacies, big-box stores, and online. Examples include self-monitoring blood glucose test systems and at-home pregnancy tests. Typically used outside of health care settings, these devices must have (1) low potential for misuse; (2) benefits that outweigh their safety risks; and (3) sufficient labeling so that lay-users can self-diagnose their condition, self-determine that the device is appropriate for their condition, and understand how to use the device without clinician assistance or instruction [10]. To minimize use errors and their associated risks, the FDA requires premarket review of the safety and effectiveness for most direct-to-consumer medical devices. This review process includes rigorous human factors usability testing with representative users to validate that the warnings, cautions, and contraindications are transparent, assess if users can complete critical tasks of the device, and identify areas of foreseeable misuse (including use of the device by unintended users) [11].

Numerous studies have evaluated the safety and usability of clinician-facing medical devices (eg, smart intravenous infusion pumps) [12-15], as well as patient-facing medical devices (eg, glucometers, ventilators, and epinephrine injectors) [16]. Typically, clinicians still mediate the use of these patient-facing medical devices either by providing a prescription to use the device or instructions on how to use the device appropriately [17]. There remains limited research on the safety and usability of direct-to-consumer medical devices, especially those used by underserved populations (eg, people with hearing impairment) [18-20].

Compared to clinician-facing medical devices, patient-facing devices present a challenge to designers as the potential environment of use and characteristics of end users is much more diverse [16]. The usability of direct-to-consumer devices is especially important given the lack of clinical supervision to support appropriate use. For instance, over-the-counter hearing aids will be on the shelf at consumer electronics stores next to other, nonmedical devices (eg, wireless headphones) and consumers may come across the device without previously identifying a need for a hearing aid. While the risk is minimal, use of hearing aids by those with normal hearing could result in noise-induced hearing loss. In this context, the packaging of the device and communication of its intended use (and contraindications to use such as an ear infection, tinnitus, or severe hearing loss) is especially important. In this study, we evaluated the safety and usability of an over-the-counter hearing aid among adults in the United States.

Methods

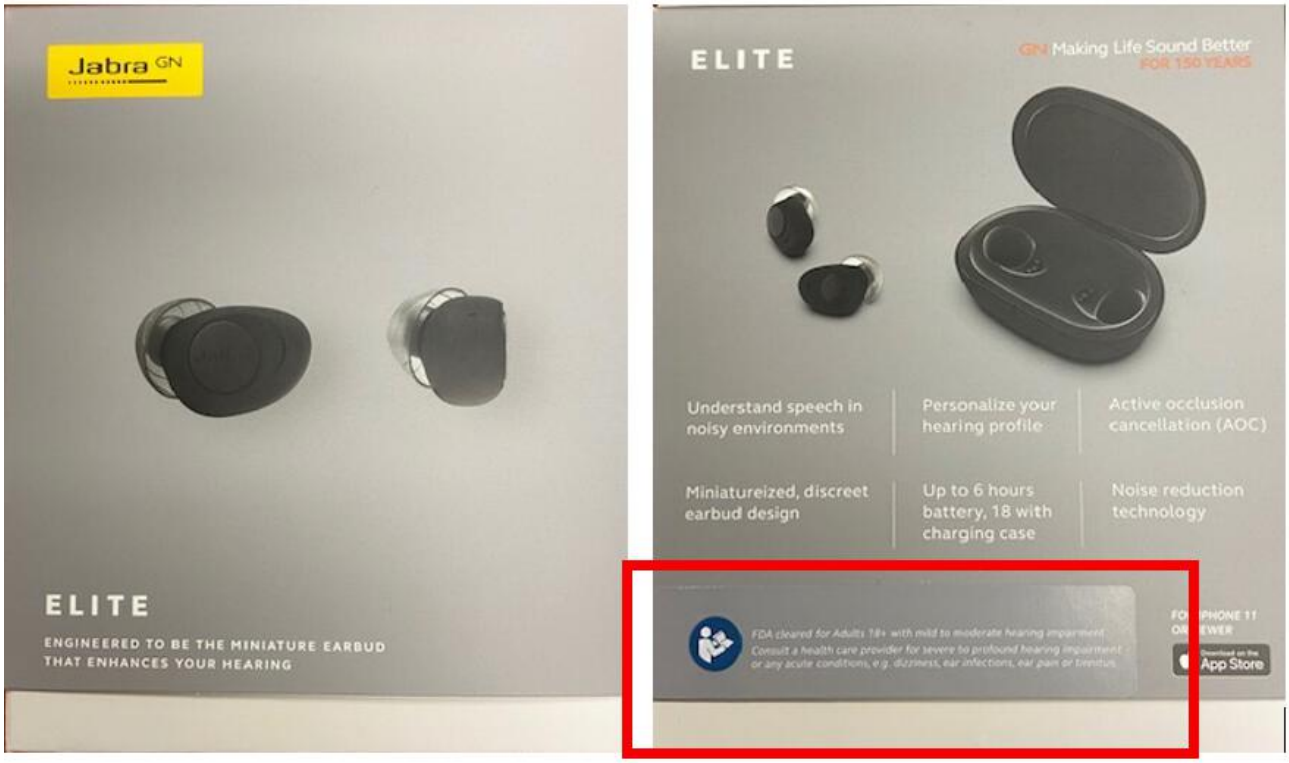
Ethical Considerations

We conducted usability testing of an over-the-counter hearing aid to evaluate the safety and usability of the device prior to market release. This study was approved by the Vanderbilt University Medical Center's institutional review board (201894). At the start of each session, we reviewed the consent document with participants and answered any questions they had. All participants provided written consent to be in the study. Data have been deidentified.

Hearing Aid

We evaluated the Jabra Enhance Plus (formerly called the "Jabra Elite" during the study), an over-the-counter hearing aid that is intended to enhance hearing but can also be used for phone calls and listening to music. Users can purchase these hearing aids at a suggested retail price of US $799 per pair. The hearing aids are physically similar to wireless earbuds with adjustable ear gels. Users can pair them to a smartphone and personalize their own hearing aid profile using a free smartphone app (ie, Jabra Enhance app). The miniaturized in-the-ear style earbuds (see Figure 1) come in a small charging case and can be used up to 12 hours on a single charge.

Figure 1. Front and back of the device box.



Participants and Setting

We conducted usability testing with production equivalent prototype devices (ie, packaging, hearing aids, charging case, and user manual) including an iPhone with the Jabra Enhance app installed. We only used iPhones in the usability testing since the Jabra Enhance app was only available for iPhones at the time of testing. We recruited two participant populations for the study, (1) intended users of the device and (2) nonintended users (Textbox 1). We identified potential participants meeting our study criteria through medical record chart review, university newsletters, and word of mouth. Eligible participants conducted a prescreening questionnaire, via phone call or email, to verify they met the inclusion criteria for the study before scheduling. The testing occurred from December 2020 to August 2021.

Textbox 1. Eligibility criteria for intended and nonintended users.

Intended users (adults aged 18+) with:

- Mild to moderate hearing impairment

Nonintended users (adults aged 18+) with:

- Normal hearing,
- Severe to profound hearing impairment, and/or
- Acute conditions including tinnitus, dizziness, ear infections, and ear pain

Intended Users

To start, we conducted usability testing with 5 intended users of the device. Based on the safety and usability issues identified in the testing round (round 1), we made recommendations for modifications to the device packaging and app screens. We then conducted round 2 of usability testing of the revised product with 15 intended users. In each usability testing session, participants were asked to complete 5 tasks (see Textbox 2) within 90 minutes. Participants received a US $50 gift card at the end of the session.

Textbox 2. Task descriptions.

Tasks for intended users (rounds 1 and 2)

- Read outer label and indicate if use is appropriate
- Set up the hearing aids and app including
- Find and open hearing aid app
- Register product and accept terms and conditions
- Pair the hearing aids
- Finalize the set up of the hearing aids
- Fit the hearing aids into ear
- Recharge the hearing aids
- Clean the hearing aids

Tasks for nonintended users (round 3)

- Read outer label and indicate if use is appropriate
- Read the inner box contents (note that the inner box contents were available to the participants, but they did not receive explicit instructions on how to read them) and indicate if use is appropriate

Nonintended Users

We conducted a third round of usability testing with 21 nonintended users of the devices to evaluate if consumers could ascertain that they should not use the device based on the external and internal labeling, user manual, and app screens. We recruited participants who were contraindicated to use the device including adults with normal hearing, adults with severe to profound hearing loss, and adults with tinnitus or severe dizziness. Round 3 participants were scheduled for a 30-minute test session, in which they were asked to complete 2 tasks (see Textbox 2). Participants received a US $20 gift card at the end of the session.

Data Collection

We conducted functional simulated use studies with end users completing real-world tasks. One human factors engineer facilitated the sessions. In line with real-world use of the device, participants received no training prior to testing. We read task instructions out loud while displaying the same instructions on a computer monitor with closed captioning. Participants were directed to complete 5 (rounds 1 and 2) or 2 tasks (round 3, see Textbox 2). After each task, participants rated how easy or difficult the task was to complete on a scale from 1 (very difficult) to 5 (very easy). After all tasks were completed, we asked intended and nonintended users 10 questions to assess their knowledge of the device and its contraindications for use (see Multimedia Appendix 1). Participants then completed the System Usability Scale (SUS), a 10-item validated survey for assessing the usability of interactive systems [21] (see Multimedia Appendix 2). To conclude the session, we conducted debrief interviews to gather additional qualitative feedback about the safety and usability of the device. The interview guide can be found in Multimedia Appendix 3. All sessions were video recorded and uploaded to Morae (TechSmith), a software suite for usability testing that supports video annotation (eg, of use errors) and logging of tasks.

Data Analysis

One researcher coded the videos in Morae, annotating areas of difficulty, confusion, safety-related issues, and errors. We analyzed the participant's path (ie, steps) through the attempted completion of each task. Any deviation that occurred in the attempt to complete the task (eg, participant clicks on an incorrect menu item or participant struggled to get the user manual out of the box) was coded as a "use error." For task 2, we specifically noted any instances where participants' workflow deviated from the expected workflow (see steps in Textbox 2). We analyzed the time spent on each task and if the tasks were (1) completed, (2) completed with difficulty, or (3) not completed (task failure). We then calculated the task success rate. A task was only scored as a success if the user was able to complete the task requirements on their own without facilitator assistance. We calculated participants' rating of difficulty for each task and the SUS scores. One researcher listened to the audio recording of each debrief interview and created a comprehensive list of feedback (eg, longer battery life) organized into two categories (1) what participants liked about the device and (2) areas for improvement. A team of human factors researchers reviewed all identified errors and participant difficulties completing tasks to develop design recommendations to improve the device, app, and packaging prior to market release. We mapped each design recommendation to specific human factors design principles, using the list of design heuristics described in Barton et al [22], which combined the commonly used heuristics for medical device design [23], interactive systems [24], and medical documents [25].

Results

Overview

Table 1 shows the participant demographics for each round of usability testing. Table 2 depicts the task success rates, number

of use errors, task durations, and difficulty ratings for each task and round of usability testing.

While we identified minimal safety concerns with the device, we identified numerous usability issues, which led to several design modifications. We outline the key issues identified from the usability testing in the following sections.

Table . Participant demographics for each usability testing round.

	Round 1 (intended users, n=5)	Round 2 (intended users, n=15)	Round 3 (nonintended users, n=21)
Hearing loss, n (%)			
None or normal hearing	0 (0)	0 (0)	18 (86)
Mild	2 (40)	7 (47)	0 (0)
Mild to moderate	1 (20)	2 (13)	0 (0)
Moderate	2 (40)	6 (40)	0 (0)
Severe	0 (0)	0 (0)	3 (14)
Tinnitus (ringing in ears), n (%)	0 (0)	0 (0)	5 (24)
Average age (years), n (range)	65 (57 - 73)	66 (56 - 79)	65 (52 - 77)
Women, n (%)	2 (40)	8 (53)	13 (62)
Education, n (%)			
High school or General Educational Development	0 (0)	5 (33)	5 (24)
Associate's or trade school	1 (20)	2 (13)	4 (19)
4-year college	2 (40)	3 (20)	7 (33)
Graduate or professional school	2 (40)	5 (33)	5 (24)

Table . Usability testing performance.

Measure and round	Task 1	Task 2	Task 3	Task 4	Task 5
Task success rate, n (%)					
Round 1	3 (60)	2 (40)	2 (40)	5 (100)	3 (60)
Round 2	12 (80)	10 (67)	7 (47)	15 (100)	10 (67)
Round 3	10 (48)	10 (48)	—[a]	—	—
Average task duration (minutes), mean (SD) [b]					
Round 1	0.72 (0.3)	9.33 (0.5)	1.37 (0.6)	0.50 (0.2)	1.08 (1.1)
Round 2	0.83 (0.5)	11.24 (4.0)	2.21 (0.9)	0.67 (0.6)	0.95 (0.8)
Round 3	—	—	—	—	—
Average difficulty rating, mean (SD) [c]					
Round 1	4.6 (0.5)	3.4 (1.2)	4.2 (0.8)	5 (0)	5 (0)
Round 2	4.87 (0.4)	3.27 (1.4)	3.73 (1.2)	4.53 (0.8)	4.67 (0.6)
Round 3	4.71 (0.6)	4.43 (0.9)	—	—	—
Number of use errors per participant, n					
Round 1	0	4	0	0	0
Round 2	0	4.9	0.5	0.6	0
Round 3	0	0.7	—	—	—

[a]Not applicable.

[b]Only includes times for participants who successfully completed the task.

[c]Scale from 1 to 5 (1=very difficult, 5=very easy).

Package Labeling

Overall, we found minimal safety issues with the set up and use of the device across all 3 rounds of testing. The one issue that presented potential, albeit minimal, concerns to safety was related to the accessibility of intended use and contraindications to use of the device. We found that the font size and color on the outside of the box made the intended use of the device hard to see and read (right side of Figure 1). Few participants saw or read this information on the box, and several participants indicated that they would not know the device was a hearing aid based on the outside of the box, stating "to know they are hearing aids, off the bat, I would have walked right by them."

A total of 3 participants in usability testing rounds 1 and 2 failed task 1 (read outer label) since they could not determine if it was appropriate for them to use the product based on the outside of the box. We identified the same problem with nonintended users of the device, with over half of the participants in round 3 stating that they did not know that the device was a hearing aid from the labeling on the outside of the box. All 3 of the participants with severe hearing loss and 3/4 of participants with tinnitus said (incorrectly) that the device would be appropriate for them to use based on the labeling on the outside of the box.

Device Pairing

The primary usability challenge with the device was setting up and pairing the hearing aids to the iPhone (task 2). In round 1, a total of 3 of 5 participants failed this task, and a fourth participant had difficulty completing it. A common challenge was that participants did not know that they needed to return to the app to complete the hearing aid set up after pairing the hearing aids to the iPhone. Rather, participants thought they were done setting up the device once they followed the Quick Guide pairing instructions (see Figure 2). The only participant to complete the task easily followed the workflow: open the Jabra Enhance app → pair the hearing aids using the iPhone → return to the app to complete the set up. As a result, we recommended design changes to the Quick Guide prior to round 2. As shown in Figure 3, the Quick Guide was revised to inform users to first go to the Jabra Enhance app before pairing the hearing aids to the phone. We further emphasized that users should return to the Jabra Enhance app after pairing to finish set up.

Figure 2. Quick guide instructions (December 2020 version—round 1 testing).

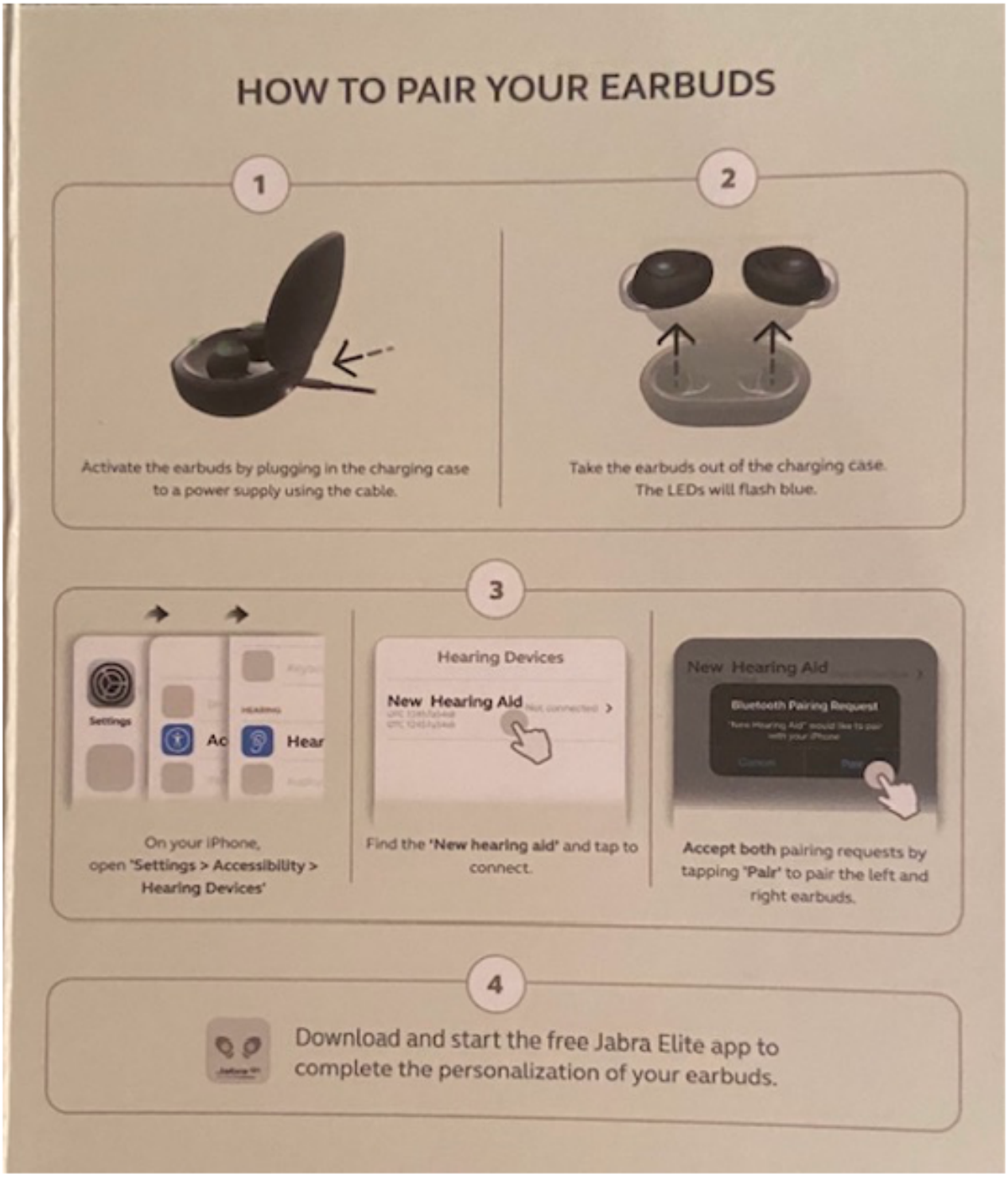
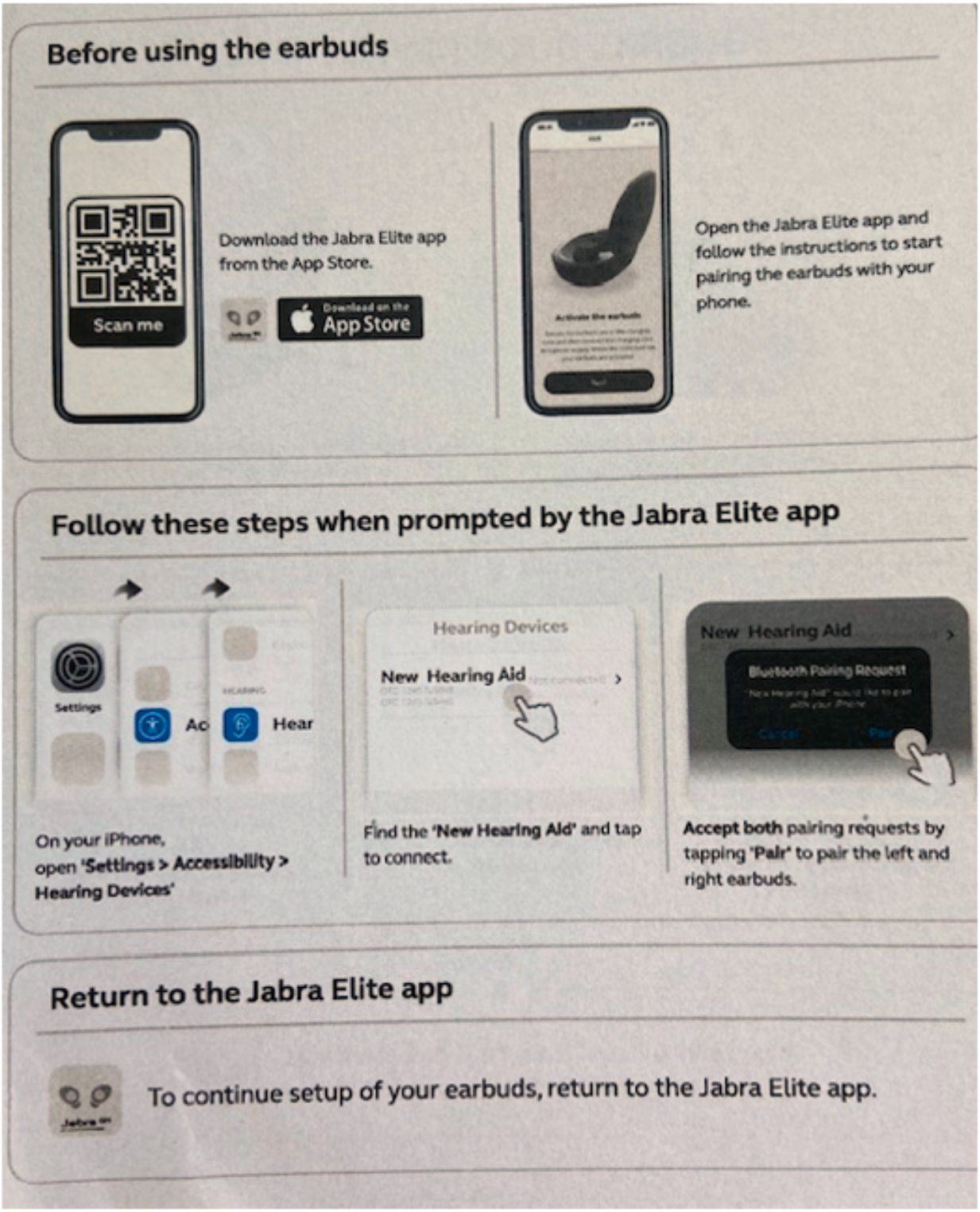


Figure 3. Quick guide (February 2021 version—round 2 testing).



Following these design changes, task 2 success rate increased from 40% (2 out of 5 in round 1) to 67% (10 out of 15 in round 2). Yet, there were still several use errors during pairing and setting up the earbuds in round 2. For instance, 2 (13%) participants did not understand that they needed to go to an app to set up the earbuds. Several participants had trouble reading the Quick Guide instructions as the text was too small (see Figure 3). A total of 3 participants did not know what to do after pairing the earbuds to the phone. Finally, it was unclear to participants which section of the iPhone settings to use to pair the hearing aids. Some participants had difficulty finding the "Accessibility" section in the iPhone's system settings, and some participants then went to the "Bluetooth" rather than the "Hearing Devices" section to pair the hearing aids.

Distinguishing the Right and Left Earbud

Participants had difficulty determining which earbud was for the right versus the left ear. Several participants stated that they

saw a "R" and "L" on the charging case, but not on the earbuds themselves. For instance, one participant stated, "so, there's not a right and a left?" while another participant said, "I don't see a left and right, they don't indicate that right?" Some participants were eventually able to identify the "R" and "L" markings on the earbuds stating,

> *Ah, there's an R on the end, that's really hard to see. I'd make that bigger...yeah that's really hard to see.*

The inability to distinguish the right from the left earbud resulted in several use errors, including difficulty placing the earbuds in the ears and difficulty placing the earbuds back in the charging case.

User Manual

The user manual contained information about the intended use and about most of the hazards of device use. The location of the user manual at the bottom of the box led to several use errors. Several participants were not able to find the manual. Finding the user manual was challenging since the text "User manual" was partially covered by the cardboard insert (see Figure 4). One participant said,

> *it's kind of hidden though...that's a little deceiving maybe, because I would have looked at this first.*

Another participant asked, "Where are the directions?" In total, 5 (33%) participants, all of whom identified the user manual, struggled to get it out of the box. Several more participants never took the manual out of the box; it is possible that these participants could not find the manual but did not verbalize this. Additionally, in round 1 testing, we found that there was not an entry in the table of contents to indicate on which page the "intended use" and "health and safety information" was within the user manual. This was corrected prior to round 2.

Figure 4. User manual cutout in box.

Product Design

Overall, participants were very positive about the device. They liked the size and design of the charging case (eg, small and easy to put in a pocket or purse), and they liked that the device was rechargeable, and thus did not require batteries. Participants liked the sleek and modern looking design stating, "it feels high-end" and said that the hearing aids were easy to place in their ears. They appreciated the different sized ear gels and the magnets that assured that the earbuds clicked into the correct charging position in the case. Participants were excited that they could purchase the device "off-the-shelf" and stated it would be helpful to use when they were in crowded rooms. Despite observed difficulties with setting up and pairing the hearing aids to the iPhone (task 2), participants reported that the app was easy to use. Finally, participants liked that they could adjust the hearing aids' settings on their phone.

Participants recommended several areas for improvement. Most participants desired a longer battery life (which was 6 h at the time of testing) so that they could wear the hearing aids all day without needing to recharge them. They also recommended that technical terms be better explained in the manual and on the box. One participant recommended that the hearing aids come in more discreet colors (ie, skin colored) instead of black. Participants recommended some modifications to the front of the box saying,

> *it should have more on the front about like if it's better for a noisier environment or something, because I'd be more likely to use it if it said, "will improve your hearing in a crowded environment" because I have a very hard time hearing when there is background noise.*

Some participants also found the contrast of white and red text on a gray background hard to read.

Knowledge Assessment

Table 3 details the scores for each of the 10 knowledge assessment questions across all 3 rounds of testing. Participants who used the user manual performed better overall compared to those who did not access or refer to the user manual. The user manual appeared to help the most on question 9—whether someone who had a recent episode of dizziness should use the earbuds.

Table . Knowledge assessment (percent correct answers).

#	Question topic	Round 1 (n=5), n (%)[a]	Round 2 (n=15), n (%)[b]	Round 3 (n=21), n (%)[c]
1	Tinnitus	1 (20)	13 (87)	13 (62)
2	Sinus infection	0 (0)	15 (100)	15 (71)
3	Severe hearing loss	4 (80)	8 (53)	5 (24)
4	Rapid hearing loss	1 (20)	11 (73)	11 (52)
5	Preference for earbuds	1 (20)	12 (80)	15 (71)
6	Laryngitis	3 (60)	4 (27)	7 (33)
7	Sore in ear	5 (100)	13 (87)	17 (81)
8	Ear wax	3 (60)	12 (80)	11 (52)
9	Dizziness	4 (80)	9 (60)	11 (52)
10	Ear infection	5 (100)	15 (100)	21 (100)

[a]Average percent correct: 50%.

[b]Average percent correct: 75%.

[c]Average percent correct: 60%.

Design Recommendations

Based on the identified usability issues, we recommended 9 design changes to the earbud system and its packaging (Table 4). We rated these recommendations as "high,""medium," or "low" priority. A total of 4 of the proposed changes (items 1, 2, 3, and 7 in Table 4) were implemented prior to subsequent testing rounds and all but 2 (items 4 and 5) were implemented prior to initial product release. Each design recommendation corresponds to a human factor's design principle (descriptions of the design principles can be found in Barton et al [22]).

Table . Design recommendations derived from the results of 3 rounds of usability testing.

#	Priority	Addressed before round 2 testing?	Addressed before product release?	Design recommendation	Corresponding human factors principle
1	High	Yes	Yes	• Increase the font size of the text in the Quick Guide	Readability (font and capitalization)
2	High	Yes	Yes	• Revise the Quick Set-up Guide and user manual to clearly show the following steps: • Download Jabra Enhance app from app store • Open the Jabra Enhance app (show a picture of the app) • Register and set up the hearing aids following the steps on the app • Pair the hearing aids with the iPhone (show pictures to show the steps of going to settings > accessibility > hearing devices), select "New Hearing Aid" device, accept both pairing requests • Return to the Jabra Enhance app (show picture of returning to app)	Organization (order)
3	High	Yes	Yes	• Add a section in the user manual's table of contents for "Intended use and warnings"	Organization (navigation tools)
4	High	No	No	• Add a screen in the app that lists the contraindications of use (eg, severe hearing loss, tinnitus, and dizziness). This could be a screen after the registration in the app that participants acknowledge they understand reasons they should not use the device.	Content (emphasis)

#	Priority	Addressed before round 2 testing?	Addressed before product release?	Design recommendation	Corresponding human factors principle
5	Medium	No	No	• Increase the font size of text about contraindications and intended use on the outside of the box.	Readability (font and capitalization)
6	Medium	No	Yes	• More clearly indicate on the box that the device is to support those with hearing loss. It may be helpful to add the below text from the user manual, page 1, somewhere onto the outside of the box: "This product may help you, if you: • Strain to follow conversations in both quiet and noisier environments • Miss important information during conversations • Have trouble hearing at a distance • Have trouble understanding the television or telephone calls"	Content (clarity of content)
7	Low	Yes	Yes	• Revise the app name and other aspects in the app and user materials to have a consistent term for the Jabra earbuds, for example, "Jabra Elite," "Jabra earbuds," and "Jabra hearing aid."	Comprehensibility (terminology)
8	Low	No	Yes	• Make the cut-out in the white box bigger so that the entire "User Manual" text can be seen (Figure 4), and add text directing users that the user manual is under the white box. Could say "User Manual" with an arrow pointing below the box.	Readability (layout and position)
9	Low	No	Yes		

#	Priority	Addressed before round 2 testing?	Addressed before product release?	Design recommendation	Corresponding human factors principle
				• Add a note on the "manual pairing" page of the user manual that participants should refer to the "first time use" section if it is their first time setting up the device: "if this is the first-time use, please refer to page X."	Organization (navigational tools)

Device Packaging

We recommended that the warnings and intended use information on the outside of the box be better emphasized (eg, larger font size and more color contrast) to ensure appropriate use. Due to participants' difficulty finding the user manual and getting it out of the bottom of the box, we recommended placing the user manual on the top so the user sees and touches it before accessing the product. This could improve safety as the user manual contains most of the warnings and intended use information about the device.

Pairing the Device

To better support the pairing of the device to the smartphone, we recommended that the Quick Guide be revised. Despite modifications after round 1 testing, participants still could not read or follow the Quick Guide instructions due to the small text size. We recommended that the font size in the Quick Guide be significantly increased given the older population intended to use this device. We also recommended modifying the Quick Guide to clearly outline the pairing steps (eg, go to "Settings" then "Accessibility" then "Hearing devices"). As participants did not know to return to the app to complete set up after pairing the earbuds to the iPhone, we recommended that the steps in the Quick Guide be more clearly sequentially marked (eg, 1, 2, 3,... or a, b, c,...).

Discussion

Principal Findings

In this study, we evaluated the safety and usability of an over-the-counter hearing aid for people with mild to moderate hearing loss. We conducted usability testing with 20 intended and 21 nonintended users of the device. We identified numerous usability issues hindering the optimal use of the hearing aid, and generated design recommendations to mitigate use errors and improve the overall design of the device.

This work expands our understanding of the safety and usability of novel, over-the-counter hearing devices [18,26]. Overall, we found minimal safety-related issues with the device. The primary safety risk is that the diagnosis and treatment of a serious condition, such as sudden deafness due to infection, may be delayed due to the lack of medical screening prior to the device use [27]. Consistent with this, the main challenge we found was related to the lack of understanding by nonintended users, including those with tinnitus and severe hearing loss, that they should not use the device. This issue underscores the importance of ensuring that device contraindications and safety warnings are clearly displayed and legible, preferably on the outside packaging. Further, redundant messaging should be incorporated; we recommended a verification dialog in the phone app during device set up to ensure that users understood device use contraindications.

Usability testing identified numerous usability issues with the device, which largely related to violations of human factors design principles. Some examples of violations of the principle, readability (font and capitalization), were the small font size of the contraindications on the outside of the box, the Quick Guide, and the markings of "L" and "R" on the hearing aids. We also identified issues with comprehensibility (terminology), as the device was referred to by different terms (eg, "Elite" and "Rogue") throughout the packaging and app. Based on our design recommendations, there were significant improvements in the usability of the device between the testing rounds. Human factors design principles [23,24], including those developed specifically for patient-facing medical documents [22,25], should be incorporated into the design of patient-facing medical devices. In hearing care, specific design guidelines for older adults (eg, use 12-14 pt font sizes, write in third-sixth grade level) [28] should be followed to maximize the ease of use of these devices given the substantial diversity in patients' physical and cognitive attributes and abilities, experience, and background.

We observed appreciable tension between the goals of designing for safety and designing for ease of use, aesthetics, and marketability. For instance, from a safety perspective, the device's contraindications for use (eg, tinnitus) should have been displayed in a large font such that they covered much of the back of the box. However, this would not be aesthetically appealing and there were concerns that it would deter customers from purchasing the device. Similarly, from a safety perspective, we recommended adding a hard stop in the app setup to ensure that the user understood the intended use of the device. However, this would have decreased the ease of pairing the earbuds to the phone. With over-the-counter medical devices, both product safety and aesthetic appeal are important [16] and

it can be challenging to achieve both concurrently resulting in trade-off decisions. For devices with greater safety risks, safety should be prioritized.

We found that participants were excited about an over-the-counter hearing aid and appreciated this product's design attributes. Similar to prior studies [18,26], participants liked the earbuds' discreet size and form factor, which made hearing aid use less obvious, especially given the growing general use of earbuds for listening to music or talking on the phone. Participants also liked that the device did not require batteries and that it had a small carrying case that could easily fit in a pocket or purse. Our study identified excitement about the device from the intended user population, which may result in increased access to and use of hearing aids.

We uncovered challenges in conducting usability testing with nonintended users of the device. One goal of the FDA device approval process is to identify areas of foreseeable misuse including use of the device by unintended users [11]. This is a unique issue with over-the-counter devices, as there is no clinical mediation to ensure appropriate use. During testing, we found that only 50% of nonintended users were able to determine that they should not use the hearing device. While the design of the device and packaging likely contributed to this, another challenge is that participants may have assumed they should use the device since they were recruited to participate in the study. In essence, participants may have been "tricked" into thinking they should use the device due to the nature of usability testing. Future research should explore additional methods or approaches for eliciting information from nonintended users of over-the-counter medical devices.

Another limitation of this study is that the findings are based on users' interaction with the device over only a short period of time; it is possible that as users learn to use the device, ease of use would increase or alternatively, new usability issues may emerge over prolonged use that we did not observe. We only used iPhones for testing and some of the participants were Android users. Despite diversity in participants' age, hearing ability, comfort with technology, and education level, our sample may not have included the full range of end user cognitive abilities, experiences, and backgrounds. In the real environment of use, older adults may rely on adult children or other family members to assist with device pairing and set up; our study highlights the usability issues encountered by older adults setting up the device on their own, without assistance. We also did not test actual use of the hearing aid to enhance participant hearing so we cannot draw conclusions on the function of the hearing aid technology. Future studies should evaluate the use of the hearing aid in the real environment over time to determine if it supports users' needs. There is increasing evidence, however, demonstrating efficacy of over-the-counter hearing aids in comparison to prescriptive devices [29,30].

Conclusions

This study describes the safety and usability of an over-the-counter hearing aid for adults with mild to moderate hearing loss. Based on the usability testing, we proposed human factors design recommendations to enhance the usability and safety of the device. For instance, the intended user group and contraindications for use (eg, tinnitus) should be clearly displayed on the outside of the box, as well as throughout the set up materials. Clear step-by-step guidelines should be available to support user set up and pairing of the device. We identified challenges to the design and testing of direct-to-consumer devices. As the population ages and technologies continue to pervade every aspect of our lives, the direct use of medical devices by laypersons will continue to expand. This study lays the foundation for future studies on best practices for the user interface design of direct-to-consumer medical devices.

Acknowledgments

This research was supported in part by GN Group (Ballerup, Denmark). Some of MES's effort in preparing this study was supported by the National Library of Medicine Institutional Training Program in Biomedical Informatics and Data Science through the National Institutes of Health (NIH; grant T15LM007450-19) and the Agency for Healthcare Research and Quality (AHRQ; grants K12HS026395 and K01HS029042). The content is solely the responsibility of the authors and does not necessarily represent the official views of the GN Group, NIH, or AHRQ.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Knowledge assessment.
[DOCX File, 17 KB - humanfactors_v12i1e65142_app1.docx]

Multimedia Appendix 2
System Usability Scale.
[DOCX File, 20 KB - humanfactors_v12i1e65142_app2.docx]

Multimedia Appendix 3

Interview questions.
[DOCX File, 19 KB - humanfactors_v12i1e65142_app3.docx]

References

1. Goman AM, Lin FR. Prevalence of hearing loss by severity in the United States. Am J Public Health 2016 Oct;106(10):1820-1822. [doi: 10.2105/AJPH.2016.303299] [Medline: 27552261]
2. Lin FR, Thorpe R, Gordon-Salant S, Ferrucci L. Hearing loss prevalence and risk factors among older adults in the United States. J Gerontol A Biol Sci Med Sci 2011 May;66(5):582-590. [doi: 10.1093/gerona/glr002] [Medline: 21357188]
3. Lin FR, Ferrucci L. Hearing loss and falls among older adults in the United States. Arch Intern Med 2012 Feb 27;172(4):369-371. [doi: 10.1001/archinternmed.2011.728] [Medline: 22371929]
4. Lin FR, Yaffe K, Xia J, et al. Hearing loss and cognitive decline in older adults. JAMA Intern Med 2013 Feb 25;173(4):293-299. [doi: 10.1001/jamainternmed.2013.1868] [Medline: 23337978]
5. Thomson RS, Auduong P, Miller AT, Gurgel RK. Hearing loss as a risk factor for dementia: a systematic review. Laryngoscope Investig Otolaryngol 2017 Apr;2(2):69-79. [doi: 10.1002/lio2.65] [Medline: 28894825]
6. Reed NS, Garcia-Morales EE, Myers C, et al. Prevalence of hearing loss and hearing aid use among US Medicare beneficiaries aged 71 years and older. JAMA Netw Open 2023 Jul 3;6(7):e2326320. [doi: 10.1001/jamanetworkopen.2023.26320] [Medline: 37505496]
7. Chien W, Lin FR. Prevalence of hearing aid use among older adults in the United States. Arch Intern Med 2012 Feb 13;172(3):292-293. [doi: 10.1001/archinternmed.2011.1408] [Medline: 22332170]
8. Hearing Health Care for Adults: Priorities for Improving Access and Affordability: National Academies Press; 2016.
9. Food and Drug Administration, Department of Health and Human Services (HHS). Medical devices; ear, nose, and throat devices; establishing over-the-counter hearing aids. Fed Regist 2022 Aug 17;87.
10. Over-the-counter (OTC) medical devices: considerations for device manufacturers. U.S. Food & Drug Administration. URL: https://www.fda.gov/medical-devices/products-and-medical-procedures/over-counter-otc-medical-devices-considerations-device-manufacturers [accessed 2023-08-10]
11. Applying human factors and usability engineering to medical devices. : U.S. Department of Health and Human Services Office of Device Evaluation; 2016.
12. Weinger MB, Wiklund ME, Gardner-Bonneau DJ. Handbook of Human Factors in Medical Device Design: CRC Press; 2010.
13. Anders S, Miller A, Joseph P, et al. Blood product positive patient identification: comparative simulation-based usability test of two commercial products. Transfusion 2011 Nov;51(11):2311-2318. [doi: 10.1111/j.1537-2995.2011.03185.x] [Medline: 21599676]
14. Harris B, Weinger MB. An insidious failure of an oxygen analyzer. Anesth Analg 2006 May;102(5):1468-1472. [doi: 10.1213/01.ane.0000205814.47841.16]
15. Weinger MB. Anesthesia equipment and human error. J Clin Monit Comput 1999 Jul;15(5):319-323. [doi: 10.1023/a:1009919207870] [Medline: 12568138]
16. Gardner-Bonneau DJ. Home health care. In: Weinger M, Wiklund M, Gardner-Bonneau DJ, editors. Handbook of Human Factors in Medical Device Design: FL: CRC Press; 2011:747-770.
17. Kingsley PA, Backinger CL, Brady MW. Medical-device user instructions: the patient's need, the nurse's role. Home Healthc Nurse 1995;13(1):31-37. [Medline: 7860339]
18. Maidment DW, Ali YHK, Ferguson MA. Applying the COM-B model to assess the usability of smartphone-connected listening devices in adults with hearing loss. J Am Acad Audiol 2019 May;30(5):417-430. [doi: 10.3766/jaaa.18061] [Medline: 31044693]
19. Story MF, Luce AC, Rempel DM. Accessibility and usability of diabetes management devices for users with vision disabilities. HFES Ann Conf Proc 2009 Oct 19;53(11):699-703. [doi: 10.1518/107118109X12524442636706]
20. Benda NC, Creber RMM. Underserved populations: integrating social determinants of health into the study of patient work. In: Valdez RS, Boca R, Holden RJ, editors. The Patient Factors: Applications of Patient Ergonimcs: CRC Press; 2021, Vol. 1:163-186. [doi: 10.1201/9780429293009-13]
21. Brooke J. SUS: a "quick and dirty" usability scale. In: Jordan PW, Werdmeester BA, McClelland AL, editors. Usability Evaluation in Industry: Taylor & Francis; 1996:189-194.
22. Barton HJ, Salwei ME, Rutkowski RA, et al. Evaluating the usability of an emergency department after visit summary: staged heuristic evaluation. JMIR Hum Factors 2023 Mar 9;10(1):e43729. [doi: 10.2196/43729] [Medline: 36892941]
23. Zhang J, Johnson TR, Patel VL, Paige DL, Kubose T. Using usability heuristics to evaluate patient safety of medical devices. J Biomed Inform 2003;36(1-2):23-30. [doi: 10.1016/s1532-0464(03)00060-1] [Medline: 14552844]
24. Scapin DL, Bastien JMC. Ergonomic criteria for evaluating the ergonomic quality of interactive systems. Behav Inf Technol 1997 Jan;16(4-5):220-231. [doi: 10.1080/0144929971198806]
25. Tremoulet P, Krishnan R, Karavite D, et al. A heuristic evaluation to assess use of after visit summaries for supporting continuity of care. Appl Clin Inform 2018 Jul;9(3):714-724. [doi: 10.1055/s-0038-1668093] [Medline: 30208496]

26. Gomez R, Habib A, Maidment DW, Ferguson MA. Smartphone-connected hearing aids enable and empower self-management of hearing loss: a qualitative interview study underpinned by the behavior change wheel. Ear Hear 2022;43(3):921-932. [doi: 10.1097/AUD.0000000000001143] [Medline: 34653030]
27. Manchaiah V, Taylor B, Dockens AL, et al. Applications of direct-to-consumer hearing devices for adults with hearing loss: a review. Clin Interv Aging 2017;12:859-871. [doi: 10.2147/CIA.S135390] [Medline: 28553093]
28. Caposecco A, Hickson L, Meyer C. Assembly and insertion of a self-fitting hearing aid: design of effective instruction materials. Trends Amplif 2011 Dec;15(4):184-195. [doi: 10.1177/1084713811430837] [Medline: 22389434]
29. Humes LE, Rogers SE, Quigley TM, Main AK, Kinney DL, Herring C. The effects of service-delivery model and purchase price on hearing-aid outcomes in older adults: a randomized double-blind placebo-controlled clinical trial. Am J Audiol 2017 Mar 1;26(1):53-79. [doi: 10.1044/2017_AJA-16-0111] [Medline: 28252160]
30. De Sousa KC, Manchaiah V, Moore DR, Graham MA, Swanepoel DW. Effectiveness of an over-the-counter self-fitting hearing aid compared with an audiologist-fitted hearing aid: a randomized clinical trial. JAMA Otolaryngol Head Neck Surg 2023 Jun 1;149(6):522-530. [doi: 10.1001/jamaoto.2023.0376] [Medline: 37052929]

Abbreviations

FDA: Food and Drug Administration
SUS: System Usability Scale

Edited by K Adapa; submitted 06.08.24; peer-reviewed by D Urbanski, M Nickbakht; revised version received 30.10.24; accepted 01.11.24; published 20.01.25.

Please cite as:
Salwei ME, Anders S, Reale C, Slagle JM, Ricketts T, Weinger MB
Evaluating the Safety and Usability of an Over-the-Counter Medical Device for Adults With Mild to Moderate Hearing Loss: Formative and Summative Usability Testing
JMIR Hum Factors 2025;12:e65142
URL: https://humanfactors.jmir.org/2025/1/e65142
doi:10.2196/65142

Predictive Factors and the Predictive Scoring System for Falls in Acute Care Inpatients: Retrospective Cohort Study

Chihiro Saito[1,2], MPH; Eiji Nakatani[2,3,4], BSP, MS, PhD; Hatoko Sasaki[2], MPH, DrPH; Naoko E Katsuki[5], MD, PhD; Masaki Tago[5], MD, PhD; Kiyoshi Harada[6], MD, PhD

[1]Department of Nursing, Shizuoka General Hospital, Shizuoka, Japan

[2]Graduate School of Public Health, Shizuoka Graduate University of Public Health, 4-27-2, Kita-ando, Aoi-ku, Shizuoka, Japan

[3]Research Support Center, Shizuoka General Hospital, Shizuoka, Japan

[4]Department of Biostatistics and Health Data Science, Graduate School of Medical Science, Nagoya City University, Nagoya, Japan

[5]Department of General Medicine, Saga University Hospital, Saga, Japan

[6]Department of Medical Safety, Shizuoka General Hospital, Shizuoka, Japan

Corresponding Author:
Eiji Nakatani, BSP, MS, PhD
Graduate School of Public Health, Shizuoka Graduate University of Public Health, 4-27-2, Kita-ando, Aoi-ku, Shizuoka, Japan

Abstract

Background: Falls in hospitalized patients are a serious problem, resulting in physical injury, secondary complications, impaired activities of daily living, prolonged hospital stays, and increased medical costs. Establishing a fall prediction scoring system to identify patients most likely to fall can help prevent falls among hospitalized patients.

Objectives: This study aimed to identify predictive factors of falls in acute care hospital patients, develop a scoring system, and evaluate its validity.

Methods: This single-center, retrospective cohort study involved patients aged 20 years or older admitted to Shizuoka General Hospital between April 2019 and September 2020. Demographic data, candidate predictors at admission, and fall occurrence reports were collected from medical records. The outcome was the time from admission to a fall requiring medical resources. Two-thirds of cases were randomly selected as the training set for analysis, and univariable and multivariable Cox regression analyses were used to identify factors affecting fall risk. We scored the fall risk based on the estimated hazard ratios (HRs) and constructed a fall prediction scoring system. The remaining one-third of cases was used as the test set to evaluate the predictive performance of the new scoring system.

Results: A total of 13,725 individuals were included. During the study period, 2.4% (326/13,725) of patients experienced a fall. In the training dataset (n=9150), Cox regression analysis identified sex (male: HR 1.60, 95% CI 1.21 - 2.13), age (65 to <80 years: HR 2.26, 95% CI 1.48 - 3.44; ≥80 years: HR 2.50, 95% CI 1.60 - 3.92 vs 20-<65 years), BMI (18.5 to <25 kg/m²: HR 1.36, 95% CI 0.94 - 1.97; <18.5 kg/m²: HR 1.57, 95% CI 1.01 - 2.44 vs ≥25 kg/m²), independence degree of daily living for older adults with disabilities (bedriddenness rank A: HR 1.81, 95% CI 1.26 - 2.60; rank B: HR 2.03, 95% CI 1.31 - 3.14; rank C: HR 1.23, 95% CI 0.83 - 1.83 vs rank J), department (internal medicine: HR 1.23, 95% CI 0.92 - 1.64; emergency department: HR 1.81, 95% CI 1.26 - 2.60 vs department of surgery), and history of falls within 1 year (yes: HR 1.66, 95% CI 1.21 - 2.27) as predictors of falls. Using these factors, we developed a fall prediction scoring system categorizing patients into 3 risk groups: low risk (0-4 points), intermediate risk (5-9 points), and high risk (10-15 points). The c-index indicating predictive performance in the test set (n=4575) was 0.733 (95% CI 0.684 - 0.782).

Conclusions: We developed a new fall prediction scoring system for patients admitted to acute care hospitals by identifying predictors of falls in Japan. This system may be useful for preventive interventions in patient populations with a high likelihood of falling in acute care settings.

(JMIR Hum Factors 2025;12:e58073) doi:10.2196/58073

KEYWORDS

falls; inpatient falls; acute care hospital; predictive factor; risk factors

Introduction

In 2022, the proportion of the population aged 65 years or older in Japan reached a record high of 29% [1]. A 2017 survey by the Ministry of Health, Labour and Welfare of Japan revealed that patients aged 75 years or older accounted for 41.5% of all patients admitted to acute care hospitals [2]. Given the trend in population aging, it is projected that the number of older adult

patients with a high risk of falling will further escalate in the future. Falls are not limited to older adults, and falls in hospitalized patients can lead to severe physical injuries, secondary complications, a marked decline in activities of daily living (ADL), and even death in extreme cases [3,4]. Therefore, fall prevention has become an important issue to protect patients' lives and quality of life [5].

Interventions for fall prevention must be strategically targeted to populations with a high risk of falling during hospitalization. Furthermore, previous studies have emphasized the importance of patient exercise therapy [6,7] and education for both patients and health care providers in fall prevention [8-12]. Educating patients about the risks of falls and strategies to mitigate these risks is crucial in reducing the incidence of falls in hospitalized patients. To effectively conduct patient education, it is imperative to construct a fall prediction model for the accurate identification of these high-risk patients. Currently, fall prevention measures in hospitals include fall prediction models using information from electronic health record (EHR) systems [13-22], as well as predictive models that analyze patient information from EHRs and nursing records using artificial intelligence [15,23]. Here, we present several fall prediction models that can be used with EHRs [17-20]. The STRATIFY scale [17] uses a history of falls, visual impairment, mental status, frequency of elimination, and ability to transfer and move as factors in a prediction model. The Morse Fall Scale [18] includes 6 items related to a history of falls, comorbidities, use of walking aids, intravenous fluids, ability to walk and move, and mental status. The Medication Fall Risk Score and Evaluation Tool [19] assesses the medication-related fall risk. This tool considers a patient's use of medications as predictors, classified according to the associated risk levels. Tago et al [20] reported 8 predictors of falls in people with disabilities in Japan: age, sex, emergency hospitalization, admission to neurosurgery, use of sleeping pills, history of falls, independence in eating, and level of independence in daily living.

Although fall prediction tools are widely used in Japanese hospitals, existing models vary in predictors and are often difficult to apply due to differences in facility and patient characteristics. Many hospitals also rely on tools that lack a strong evidence base [24,25] and are not well integrated with EHR systems. In our clinical environment, we found it challenging to implement existing models due to their reliance on numerous variables that are either difficult to extract from the EHR or unavailable. Therefore, we aimed to develop a streamlined fall risk assessment tool that considers facility-specific factors and can be seamlessly integrated with EHR systems, enabling real-time insights and more efficient fall prevention strategies.

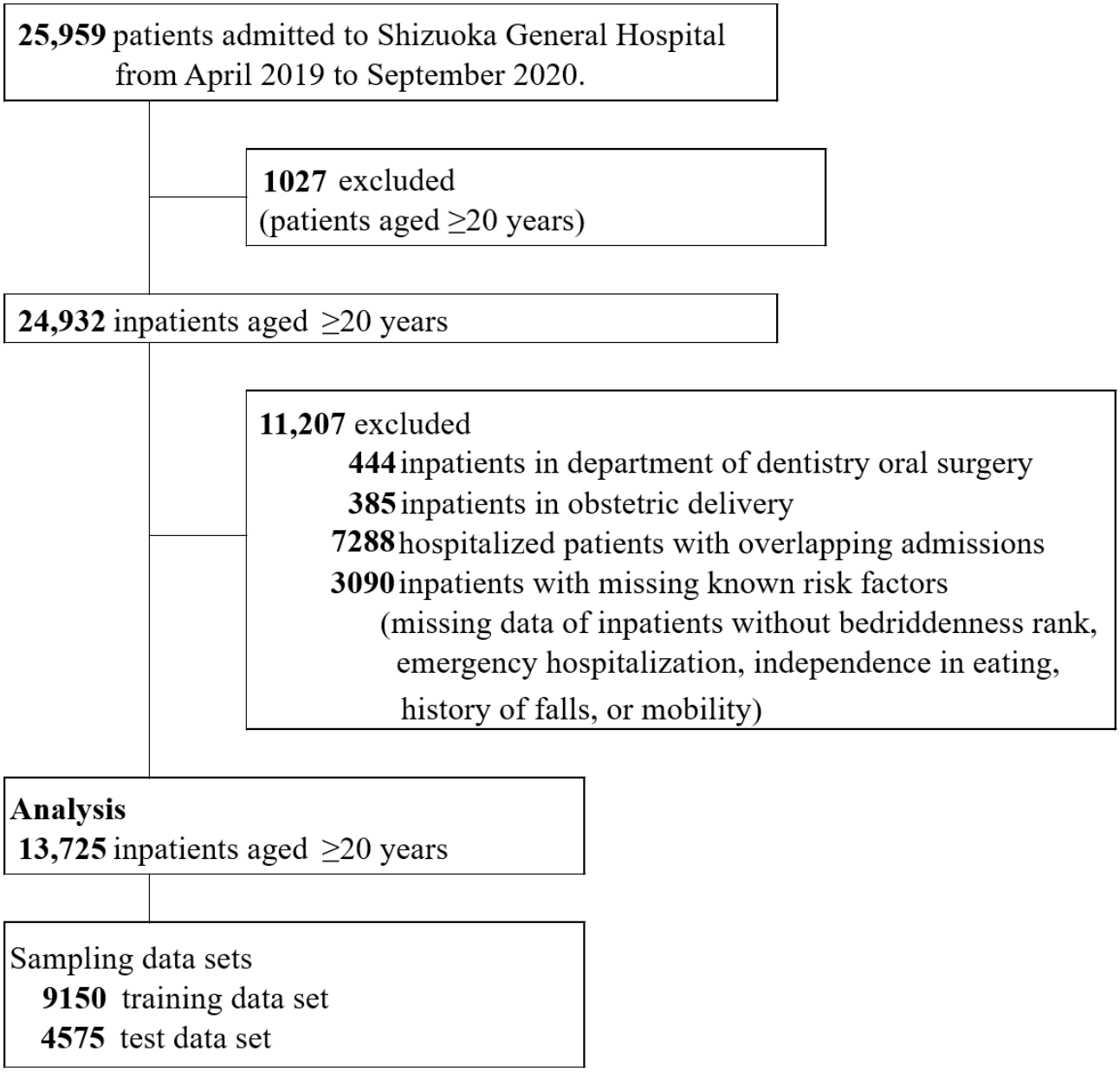
The purpose of this study was to identify predictive factors for falls in patients admitted to an acute care hospital as well as to develop a scoring system using these factors and evaluate its validity.

Methods

Participants and Study Design

We conducted a retrospective cohort study of patients aged ≥20 years admitted to Shizuoka General Hospital between April 2019 and September 2020. Inpatients excluded from the study included those not covered by the Diagnosis Procedure Combination system, such as dental and oral surgery patients and obstetrics and gynecology inpatients during pregnancy, childbirth, and postpartum. In addition, inpatients lacking data on known risk factors such as the degree of independence in daily living for the older adults with disabilities, bedriddenness rank (BR), emergency admission, dietary independence, mobility, and history of falls in the past year were also excluded, as illustrated in Figure 1.

Figure 1. Flowchart illustrating the study population.



Variables at Hospitalization as Candidate Predictive Factors

All data used in this study were extracted from patients' medical records as of February 3, 2022. Preadmission medical history variables included dementia [5,6], Parkinson disease [7,26], stroke [6], visual impairment (with or with no diagnosis of glaucoma or cataracts) [4,5], history of falls [5,16,20,27], and use of sleep medications [20,28-30]. The following variables at admission were also collected: age [6,16,20,27], sex [20,27,28], BMI [31-35], date of admission [20], disease name at admission, department [20], mode of admission, ambulance transport, consciousness disorders [16,23,28], requirement for nursing care, good sleep condition, use of sleeping medication [20,28,30], status of medication management, BR [20,36], Cognitive Function Scores [5,6], ADL at admission (eating, transferring, dressing, toilet transfer or use, bathing, level walking, stair use, changing clothes, defecation management, and urinary management) [17,18,20], fall assessment end points at admission (history of falls or falls within 1 year, inability to stand without holding on [28], impaired judgment and comprehension, toilet assistance, and use of portable toilet), and presence of physical restraint screening at admission.

The BR [20,36] is an official assessment tool in Japan's long-term care insurance system [37]. The BR is an assessment of the degree to which a person's daily life is restricted; this degree is mainly assessed in terms of mobility in daily life, such as whether the person is independent, in a wheelchair, or bed bound. The Ministry of Health, Labour and Welfare ranks the degree of BR based on evaluations by nurses and other health care professionals according to the daily care they provide during hospitalization, as well as on reports from family members. The procedure for assessing BR and its reliability have been reported [20,36].

Falls During Hospitalization as an Outcome

The primary end point was the time from the date of admission to a fall at incident level [38] 2 or higher (hereafter referred to

as “fall”), which requires medical resources. For patients who died during hospitalization and those who did not have a fall, the date of death or date of discharge, respectively, was used as the censoring date. The classification of incident levels is shown in Table S1 in Multimedia Appendix 1.

Statistical Analysis

Demographic data and potential predictors at the time of hospital admission were summarized as follows. Continuous variables are described using mean (SD) or median (range), considering the distribution type. Categorical variables are summarized as frequency (%). For comparisons between the groups with and with no fall occurrence, *t* tests were used for continuous variables and chi-squared tests for categorical variables. The Kaplan-Meier method was applied to estimate the fall rate.

We explored predictors of falls and constructed a predictive model using two-thirds of the total cases, randomly selected as the training group, with the remaining one-third serving as the validation group for the scoring system. In the training group, predictive factors were identified using the Cox proportional hazards model, and we calculated hazard ratios (HRs), 95% confidence intervals (CIs), and *P* values. Predictive candidates that were significant ($P<.05$) in the comparison of backgrounds among patients with and with no falls, along with known predictors, were included in a multivariable model. Factors with *P* value of <.2 in this model were identified as predictors of inhospital falls. Independence between explanatory variables was confirmed using an absolute value of Spearman rank correlation coefficient of >0.4. Among 2 correlated variables, 1 was chosen based on ease of collection or clinical significance; this variable was then included in the multivariable model. According to the identified predictive factors, a score was created for each HR, and these scores were summed. We used a method called “conditional inference tree analysis” to categorize patients into 3 groups based on their risk of falling. This approach works by first dividing the data into 2 groups based on their overall scores. Then, a statistical test is performed to see whether these 2 groups are significantly different, and the variable that shows the strongest difference (the one with the lowest *P* value) is used to split the groups. This process is repeated within each subgroup until no further meaningful divisions can be made or the smallest group size allowed is reached. The predictive performance of this score and the fall risk groups in the validation group were evaluated using the c-index.

The significance level of the 2-tailed test was set at .05. Missing values were not imputed in the analyses. All analyses were performed using R (version 4.1.1; The R Foundation for Statistical Computing), EZR (version 1.54; Saitama Medical Center, Jichi Medical University) [39], and IBM SPSS (version 28; IBM Corp).

Ethical Considerations

This study conformed to the Ethical Principles for Medical Research Involving Human Subjects issued by the Ministry of Health, Labour and Welfare and the Ministry of Education, Culture, Sports, Science and Technology of Japan. Following these guidelines, the Shizuoka General Hospital research ethical committee determined that individual patient informed consent was not required because we analyzed existing information in this study, and patients were given the right to refuse participation via disclosure. After obtaining committee approval (SGHIRB #2020075; January 15, 2022) and publishing the disclosure document on Shizuoka General Hospital’s website, the information for each individual was anonymized, and the analysis was conducted.

Results

Patient Background and Falls at Incident Level 2 or Higher on Admission

From April 1, 2019, to September 30, 2020, a total of 24,932 inpatients aged 20 years or older were admitted to Shizuoka General Hospital. We excluded 3.3% (829/24,932) of patients not covered by the Diagnosis Procedure Combination, 29.2% (7288/24,932) with duplicate admissions, and 12.4% (3090/24,932) with missing known risk factors of falls. Consequently, 55% (13,725/24,932) of patients were included in the analysis (Figure 1).

During the observation period, defined as the length of hospital stay (median [range]: 13 [1-271] days), 3.6% (489/13,725) of patients experienced falls across all incident levels, of which 2.4% (326/13,725) falls were classified as incident level 2 or higher. For this study, we used fall data meticulously managed by the hospital’s medical safety department, ensuring accuracy and reliability. The details are shown in Table S2 in Multimedia Appendix 1. The median age (range) of the patients included in the analysis was 66 (20-103) years, with 52.1% (7150/13,725) male patients, and the median BMI (range) was 22.8 (9.6 - 58.1) kg/m^2. Table S3 in Multimedia Appendix 1 shows the results of the comparison between the backgrounds of patients in the training and test datasets. In a univariable analysis of the training dataset (n=9150; two-thirds of the study population), we compared patient backgrounds according to the presence of falls (Table 1).

Table . Patient background on admission[a].

Variable		With no fall (n=8934)	With fall (n=216)	*P* value
Age				<.001
	Age (years), mean (SD)	65.9 (17)	75.0 (11.5)	
Age				<.001
	20 to <65 years, n (%)	3347 (37.5)	28 (13)	
	65 to <80 years, n (%)	3673 (41.1)	105 (48.6)	
	≥80 years, n (%)	1914 (21.4)	83 (38.4)	
Sex, n (%)				<.001
	Female	4287 (48)	75 (34.7)	
	Male	4647 (52)	141 (65.3)	
BMI				<.001
	BMI, mean (SD)	22.9 (4.2)	21.1 (3.9)	
BMI				<.001
	<18.5 kg/m^2, n (%)	1142 (12.8)	48 (22.2)	
	18.5 to <25 kg/m^2, n (%)	5363 (60)	132 (61.1)	
	≥25 kg/m^2, n (%)	2403 (26.9)	36 (16.7)	
	Missing, n (%)	26 (0.3)	0 (0)	
Dementia, n (%)				.927
	No	8697 (97.3)	211 (97.7)	
	Yes	237 (2.7)	5 (2.3)	
Parkinson disease, n (%)				.592
	No	8861 (99.2)	213 (98.6)	
	Yes	73 (0.8)	3 (1.4)	
Stroke, n (%)				.003
	No	8121 (90.9)	183 (84.7)	
	Yes	813 (9.1)	33 (15.3)	
Visual impairment, n (%)				.778
	No	7991 (89.4)	195 (90.3)	
	Yes	943 (10.6)	21 (9.7)	
Cognitive function score, n (%)				<.001
	No	7855 (87.9)	179 (82.9)	
	Yes	710 (7.9)	37 (17.1)	
	Missing	369 (4.1)	0 (0)	
Ambulance transport, n (%)				<.001
	No	7432 (83.2)	156 (72.2)	
	Yes	1502 (16.8)	60 (27.8)	
Emergency admission, n (%)				<.001
	Scheduled hospitalization	5528 (61.9)	74 (34.3)	
	Emergency hospitalization	3406 (38.1)	142 (65.7)	
Department, n (%)				.007
	Internal medicine	4496 (50.3)	118 (54.6)	
	Department of surgery	4025 (45.1)	80 (37)	

Variable		With no fall (n=8934)	With fall (n=216)	*P* value
	Emergency department	413 (4.6)	18 (8.3)	
Consciousness disorders, n (%)				<.001
	No	8022 (89.8)	167 (77.3)	
	Yes	912 (10.2)	49 (22.7)	
Bedriddenness rank, n (%)				<.001
	Rank J	5500 (61.6)	65 (30.1)	
	Rank A	1568 (17.6)	58 (26.9)	
	Rank B	526 (5.9)	34 (15.7)	
	Rank C	1340 (15)	59 (27.3)	
Eating, n (%)				<.001
	Independent	7067 (79.1)	107 (49.5)	
	Requiring assistance	1867 (20.9)	109 (50.5)	
Transferring, n (%)				<.001
	Independent	6579 (73.6)	89 (41.2)	
	Requiring assistance	2355 (26.4)	127 (58.8)	
Dressing, n (%)				<.001
	Independent	7065 (79.1)	112 (51.9)	
	Requiring assistance	1853 (20.7)	104 (48.1)	
	Missing	16 (0.2)	0 (0)	
Toilet transfer or use, n (%)				<.001
	Independent	6820 (76.3)	94 (43.5)	
	Requiring assistance	2099 (23.5)	121 (56)	
	Missing	15 (0.2)	1 (0.5)	
Bathing, n (%)				<.001
	Independent	6673 (74.7)	95 (44)	
	Requiring assistance	2042 (22.9)	114 (52.8)	
	Missing	219 (2.5)	7 (3.2)	
Level walking, n (%)				<.001
	Independent	6707 (75.1)	93 (43.1)	
	Requiring assistance	2107 (23.6)	122 (56.5)	
	Missing	120 (1.3)	1 (0.5)	
Stair use, n (%)				<.001
	Independent	6541 (73.2)	86 (39.8)	
	Requiring assistance	1986 (22.2)	111 (51.4)	
	Missing	407 (4.6)	19 (8.8)	
Changing clothes, n (%)				<.001
	Independent	6799 (76.1)	97 (44.9)	
	Requiring assistance	2116 (23.7)	119 (55.1)	
	Missing	19 (0.2)	0 (0)	
Defecation management, n (%)				<.001
	Independent	7254 (81.2)	127 (58.8)	
	Requiring assistance	1618 (18.1)	87 (40.3)	
	Missing	62 (0.7)	2 (0.9)	

Variable		With no fall (n=8934)	With fall (n=216)	*P* value
Urination management, n (%)				<.001
	Independent	7243 (81.1)	126 (58.3)	
	Requiring assistance	1625 (18.2)	88 (40.7)	
	Missing	66 (0.7)	2 (0.9)	
History of falls within 1 year, n (%)				<.001
	No	7830 (87.6)	153 (70.8)	
	Yes	1104 (12.4)	63 (29.2)	
Inability to stand without holding, n (%)				<.001
	No	5318 (59.5)	53 (24.5)	
	Yes	3486 (39)	159 (73.6)	
	Missing	130 (1.5)	4 (1.9)	
Impaired judgment and comprehension, n (%)				<.001
	No	7592 (85)	157 (72.7)	
	Yes	1116 (12.5)	53 (24.5)	
	Missing	226 (2.5)	6 (2.8)	
Toileting assistance, n (%)				<.001
	No	6694 (74.9)	106 (49.1)	
	Yes	1795 (20.1)	99 (45.8)	
	Missing	445 (5)	11 (5.1)	

[a]Between-group comparisons were made using *t* tests and χ^2 tests for continuous and categorical variables, respectively. The *P* value was calculated using the Wald test. Bedriddenness rank J=independent or autonomous, rank A=housebound, rank B=chair, and rank C=bedridden.

Predictors of Falls

In the training dataset, univariable Cox regression analysis compared patient backgrounds based on the presence or absence of falls and identified factors affecting the time from admission to the date of fall (Table 2, left side). Correlations between explanatory variables were checked using Spearman correlation coefficients, and independent explanatory variables were entered into a multivariable regression model. The variables were narrowed down by applying a high absolute value of the Spearman correlation coefficient (>0.4) (Table S4 in Multimedia Appendix 1). Among the 21 variables that were significant in the univariable analysis, 5 variables (age, sex, BMI, department, and history of falls within 1 year) were not correlated with each other. Sixteen variables (including 12 ADL-related variables, emergency admission, and consciousness disorders) had a correlation coefficient of 0.4 or higher. From the correlated variables, BR was ultimately chosen. This decision was influenced by the fact that 12 other ADL variables showed correlation, and BR could potentially explain physical severity, including factors such as emergency transport and impaired consciousness. The variables were chosen based on ease of collection or clinical significance.

Table . Univariable and multivariable Cox regression analysis results for fall rates in the training dataset[a].

Variable (reference) and category	Training dataset (n=9150)					
	Univariable model			Multivariable model		
	HR[b]	95% CI	*P* value	HR	95% CI	*P* value
Age						
1 year	1.03	1.02 - 1.04	<.001	—[c]	—	—
Age (20 to <65 years)						
65 to <80 years	2.71	1.79 - 4.12	<.001	2.26	1.48 - 3.44	<.001
≥80 years	3.53	2.29 - 5.43	<.001	2.50	1.60 - 3.92	—
Sex (female)						
Male	1.60	1.21 - 2.12	.001	1.60	1.21 - 2.13	.001
BMI						
1 kg/m^2	0.92	0.89 - 0.96	<.001	—	—	—
BMI (≥25 kg/m ^{2})						
18.5 to <25 kg/m^2	1.57	1.09 - 2.27	.017	1.36	0.94 - 1.97	.127
≤18.5 kg/m^2	2.01	1.30 - 3.10	.002	1.57	1.01 - 2.44	—
Ambulance transport (no)						
Yes	1.77	1.33 - 2.36	<.001	—	—	—
Emergency admission (scheduled hospitalization)						
Emergency hospitalization	1.22	0.90 - 1.64	.203	—	—	—
Stroke (no)						
Yes	1.18	0.81 - 1.72	.376	—	—	—
Bedriddenness rank (rank J)						
Rank A	2.27	1.59 - 3.23	<.001	1.81	1.26 - 2.60	.001
Rank B	2.93	1.93 - 4.45	<.001	2.03	1.31 - 3.14	—
Rank C	1.94	1.35 - 2.77	<.001	1.23	0.83 - 1.83	—
Cognitive function score (no)						
Yes	1.40	0.98 - 2.01	.063	—	—	—
Department (department of surgery)						
Internal medicine	1.36	1.02 - 1.81	.034	1.23	0.92 - 1.64	.119
Emergency department	2.16	1.29 - 3.60	.003	1.81	1.26 - 2.60	—
Consciousness disorders (no)						
Yes	1.47	1.06 - 2.03	.020	—	—	—
Eating (independent)						
Requiring assistance	2.05	1.56 - 2.69	<.001	—	—	—
Transferring (independent)						
Requiring assistance	2.12	1.60 - 2.80	<.001	—	—	—
Dressing (independent)						

Variable (reference) and category	Training dataset (n=9150)					
	Univariable model			Multivariable model		
	HR[b]	95% CI	*P* value	HR	95% CI	*P* value
Requiring assistance	1.88	1.43 - 2.48	<.001	—	—	—
Toilet transfer or use (independent)						
Requiring assistance	2.22	1.68 - 2.93	<.001	—	—	—
Bathing (independent)						
Requiring assistance	2.10	1.59 - 2.78	<.001	—	—	—
Level walking (independent)						
Requiring assistance	2.18	1.65 - 2.88	<.001	—	—	—
Stair use (independent)						
Requiring assistance	2.17	1.63 - 2.90	<.001	—	—	—
Changing clothes (independent)						
Requiring assistance	2.10	1.60 - 2.77	<.001	—	—	—
Defecation management (independent)						
Requiring assistance	1.63	1.23 - 2.15	.001	—	—	—
Urination management (independent)						
Requiring assistance	1.64	1.24 - 2.17	.001	—	—	—
History of falls within 1 year (no)						
Yes	2.01	1.50 - 2.70	<.001	1.66	1.21 - 2.27	.002
Inability to stand without holding (no)						
Yes	2.51	1.83 - 3.45	<.001	—	—	—
Impaired judgment and comprehension (no)						
Yes	1.40	1.03 - 1.92	.035	—	—	—
Toileting assistance (no)						
Yes	2.10	1.59 - 2.77	<.001	—	—	—

[a]Between-group comparisons were made using *t* tests and χ^2 tests for continuous and categorical variables, respectively. The *P* value was calculated using the Wald test. Bedriddenness rank J= independent or autonomous, rank A=housebound, rank B=chair, and rank C= bedridden.

[b]HR: hazard ratio.

[c]Not applicable.

In multivariable analysis, age (65 to <80 years: HR 2.26, 95% CI 1.48 - 3.44; ≥80 years: HR 2.50, 95% CI 1.60 - 3.92 vs 20 to <65 years), sex (male: HR 1.60, 95% CI 1.21 - 2.13), BMI (18.5 to <25 kg/m²: HR 1.36, 95% CI 0.94 - 1.97; <18.5 kg/m²: HR 1.57, 95% CI 1.01 - 2.44 vs ≥25 kg/m²), BR (rank A: HR 1.81, 95% CI 1.26 - 2.60; rank B: HR 2.03, 95% CI 1.31 - 3.14; rank C: HR 1.23, 95% CI 0.83 - 1.83 vs rank J), emergency department (internal medicine: HR 1.23, 95% CI 0.92 - 1.64; emergency department: HR 1.81, 95% CI 1.26 - 2.60 vs department of surgery), and history of falls within 1 year (yes: HR 1.66, 95% CI 1.21 - 2.27) are shown as predictors of inhospital falls (Table 2, right side).

Construction of a Fall Scoring System

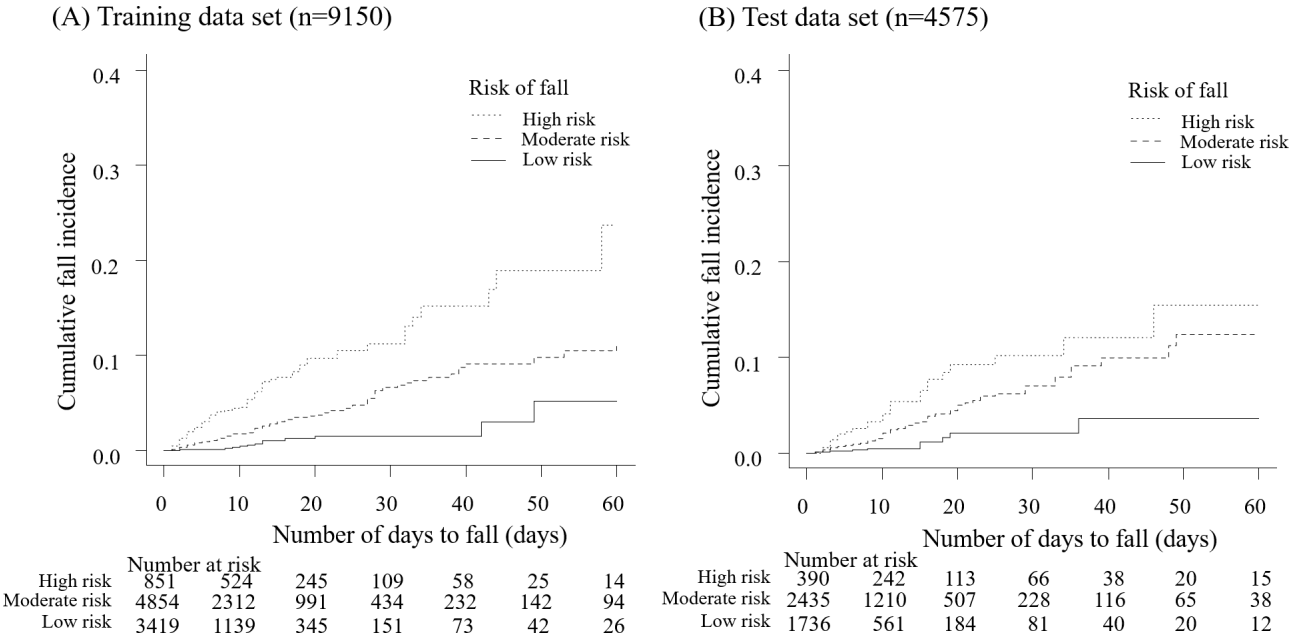
Based on the results of multivariable analysis using the training set, we weighted the scores based on each HR (Table 2, right side) and formed 3 fall-risk groups (low risk: 0-4 points, moderate risk: 5-9 points, and high risk: 10-15 points) using conditional inference tree analysis (Figure S1 in Multimedia Appendix 1). The new fall prediction scoring system built on this basis is shown in Figure 2. As a predictor of the performance of these 3 classifications of fall prediction, the c-index in the validation set (n=4561) was 0.733 (95% CI 0.684 - 0.782). The cumulative fall incidences in training and test datasets are

shown, with Kaplan-Meier curves presented for the training dataset (Figure 3A) and the test dataset (Figure 3B).

Figure 2. Fall prediction scoring system to be implemented at the time of admission.

Predictive score for falls at admission		
Sex		
☐ Female		0 points
☐ Male		2 points
Age		
☐ 20 to < 65 years		0 points
☐ 65 to < 80 years		3 points
☐ ≥ to 80 years		4 points
BMI		
☐ ≤ 18.5 kg/m^2		2 points
☐ 18.5 to < 25 kg/m^2		1 points
☐ ≥ to 25 kg/m^2		0 points
Bedriddenness rank		
☐ Rank J		0 points
☐ Rank A		2 points
☐ Rank B		3 points
☐ Rank C		1 points
Department		
☐ Internal medicine		1 points
☐ Department of surgery		0 points
☐ Emergency department		2 points
History of falls with in 1 years		
☐ Absence		0 points
☐ Presence		2 points
Sum of the points		____points
Risk of fall	0-4	Low risk
	5-9	Moderate risk
	10-15	High risk

Figure 3. Fall risk classification-specific cumulative fall incidence. Cumulative fall incidences classified by fall risk scoring are shown for the training dataset (**A**) and test dataset (**B**).



Discussion

Principal Results

This study was a retrospective cohort investigation conducted at an institution specialized in acute inpatient care, aimed at identifying the risk factors for falls using the time from admission to fall as the outcome variable. Fall risk factors included age, sex, BMI, BR, emergency department, and history of falls within 1 year. Specifically, the study found that older patients (aged 80 years and older) had a higher risk of falls, with men being more at risk than women. Patients with a BMI of <18.5 and those admitted through the emergency department had an increased risk. In addition, those with a history of falls within the past year were particularly vulnerable. We constructed a new predictive scoring system for falls by weighting scores based on each HR according to the results of multivariable analysis and using statistical methods to classify fall risk groups into 3 categories (low risk: 0-4 points, moderate risk: 5-9 points, and high risk: 10-15 points). The present fall prediction scoring system could facilitate preventive interventions for high-risk patients, potentially reducing the likelihood of falls among the most susceptible patient populations and improving patient safety and care in the hospital environment.

Comparison With Prior Work

In terms of age-specific fall incidence, it was evident that a higher proportion of falls occurred among people aged 65 years and older. This result is in alignment with previous reports [6,16,20,22,27,40] identifying advanced age as a predictive factor for falls. Consistent with past studies [5,16,20,22,40], a history of falls was determined to be a predictive factor.

Previous reports on predictors of falls have shown that sex can be a predictor for both men and women. In this study, being male was identified as a risk factor. In addition, men have been found to experience multiple falls more frequently [27]. Past reports indicating that being male increases the risk of falls [20,28] have focused on hospitalized patients whereas those suggesting an increased risk for women [41,42] have focused on community-dwelling individuals. These differences may be owing to the different characteristics of the study populations, that is, relatively healthy community residents and patients in health care facilities. This may be related to the fact that hospitalized patients tend to have reduced physical activity, which increases the risk of falls. However, the relationship between sex differences and fall risk factors in the hospitalized population remains unclear, and further research is needed to elucidate these aspects.

A systematic review targeting community-dwelling older adults showed that a low BMI (<17 kg/m^2) is associated with a greater risk of falls, when using 23.5 kg/m^2 as the baseline [34]. In addition, some reports indicate that both high (25 - 35 kg/m^2 or above) and low (below 18.5 kg/m^2) BMI values are associated with increased fall risks [31-33,35]. These findings suggest that extreme BMI values influence the risk of falling. However, in our study, we identified low BMI (below 18.5 kg/m^2) and normal BMI (18.5 - 25 kg/m^2) as risk factors for falls, using high BMI (25 kg/m^2 or above) as the reference category. This divergence in results might be attributed to the fact that in our Japanese study population, few patients met the international high BMI criterion (>30 kg/m^2). In the study, because only 5.7% (6/741) of people in this training category fell, BMI >30 kg/m^2 was not able to be included as the category for a categorical variable BMI.

The BR degree, which is strongly correlated with ADL, was identified as a risk factor for inhospital falls. This finding aligns with several fall prediction models [17,18,20], demonstrating the significant role of a decline in ADL, as it is known that exercise therapy is effective in fall prevention [6,7]. This underscores that impaired ADL is a crucial factor in determining the outcomes of patients who experience falls. In addition,

because the overall assessment of ADL is predominantly based on mobility, a substantial correlation [43] between ADL and BR has been observed. In our study, of the 15 variables that showed a correlation with BR, 12 were related to components of ADL.

Being admitted to the emergency department (as an inpatient department) was identified as a risk factor for inhospital falls. The emergency department inpatient population comprises patients who have been transported to the emergency department or who otherwise came voluntarily to the emergency department. This population typically has more severe illness, which may explain why emergency department admission is a risk factor for inhospital falls. Moreover, a history of falls within the past 1 year was identified as a risk factor for inhospital falls. This result was similar to previous reports that identified a fall history as a risk factor for subsequent falls [17,18,20].

Falls are internationally recognized as a serious health issue [3,30], and various efforts to prevent falls are undertaken worldwide. Particularly in Japan, where rapid aging is prevalent, fall prevention has become an increasingly critical issue. Education for patients and staff is considered a fundamental approach to addressing this problem [8,10-12,44], and the introduction and enhancement of educational programs in Japanese hospitals are desirable [10]. To achieve this, it is essential to identify high-priority patients among inpatients and implement fall prevention measures as a high-risk approach. Multifactorial interventions to comprehensively assess and address multiple fall risk factors have proven effective [10]. However, some tools for evaluating fall risk have been criticized for their time-consuming nature and limited effectiveness [45-47], necessitating judicious selection and effective use. Recent research supports these multifactorial interventions and highlights the importance of tailored educational programs and effective risk assessment tools. Guidelines from the United Kingdom's National Institute for Health and Care Excellence offer specific approaches for fall prevention [48], which could serve as valuable references for fall prevention strategies in Japan. In the approach to fall prevention in Japanese hospitals, international insights and guidelines should be considered while also tailoring unique approaches according to facility characteristics, health care delivery systems, and patient backgrounds.

Statistical Validation and Clinical Application of Risk Categorization

In our research, we used the conditional inference tree method for statistical analysis to categorize patients into low-, middle-, and high-risk groups, as detailed in Table S4 in Multimedia Appendix 1. This methodological choice ensured that our risk stratifications were based on solid data analysis, avoiding arbitrary determinations. In addition, this scoring approach permits adjustment of cutoff values based on each health care facility's resources and circumstances, enhancing its applicability in diverse clinical settings. Our aim is not to change the patient's fall risk level based on resources but to adjust the intensity of preventive interventions according to available resources. The fall risk classification is based solely on clinical characteristics, while resource availability guides the prioritization and distribution of these interventions.

For patients identified as high risk for falls according to our predictive model, it is important to recognize that identifying these individuals is only the first step; providing effective interventions is a separate and critical challenge. Based on prior research demonstrating their effectiveness, we recommend several interventions tailored to implement fall prevention measures as a high priority. Here, we outline interventions such as increased monitoring, personalized environmental adjustments, nonslip footwear, assistive devices such as walkers or canes, one-on-one support, and immediate assistance. In addition, patient education and rehabilitation are crucial components. Educating both patients and health care providers about fall risks and preventive strategies, combined with physical therapy to enhance strength and balance, can significantly reduce the risk of falls [6-12]. However, it may be necessary to adjust these cutoffs based on sensitivity and specificity considerations to enhance the accuracy of patient risk identification, aligning more closely with practical prediction practices in health care.

The fall prediction tool we developed integrates seamlessly with a hospital's EHR system. Upon a patient's admission, it automatically retrieves EHR data and calculates a fall risk score based on various predictors. This integration allows all health care providers, including emergency and trauma physicians, to easily access the patient's fall risk assessment. Emergency or trauma physicians can use this tool to quickly identify patients at high risk for falls and implement appropriate interventions. This proactive approach can significantly reduce the incidence of falls and enhance patient safety in fast-paced and high-stress environments such as emergency departments and trauma centers.

Limitations

This study has several limitations. First, owing to the identification of inhospital fall predictors based on the characteristics of the facility and patient population in an acute care hospital, generalizing the findings to other all hospitals may be challenging. While the generalizability of our model is not fully guaranteed due to the lack of external validation, hospitals with similar facilities and patient characteristics might find the identified predictors and scoring system developed in this study applicable to their setting. Second, our study did not examine psychiatric symptoms, including delirium, as a potential predictive factor [49,50]. Because previous studies have shown that psychiatric symptoms can be a risk factor for falls [17,18,28], it can be necessary to reconstruct the prediction model by including additional predictors of falls, including psychiatric symptoms such as delirium, in future studies. Third, we did not examine the association between medications and falls. Previous studies have shown that taking sleeping pills [19,20,28-30,51] and the number of medications [42,52-54] are risk factors for falls. At the time of our study, limitations in our EHR system made it difficult to collect accurate medication data. This prevented us from including medication type and number as predictors. Since then, our EHR system has been upgraded, and we can now reliably obtain medication data. We

plan to incorporate these variables in future studies to enhance the accuracy of our fall prediction model. Fourth, we could not compare the results of our model with those of existing fall prediction models. However, we believe that our newly developed fall prediction model is highly useful in that it is easy to apply in many Japanese hospitals with acute care settings, similar to that in this study.

To further enhance the practical application of our fall prediction scoring system, we plan to integrate it into our hospital's EHRs to gather real-world evidence. This will allow us to evaluate its usability, accuracy, and impact on reducing inhospital falls. Future research will include multisite longitudinal studies to validate the tool's effectiveness across different health care settings.

Conclusions

We successfully identified predictors of falls within a patient population admitted to an acute care hospital and developed a novel prediction model in Japan. This model could serve as an effective tool to guide preventive interventions for both individual hospitalized patients and high-risk populations in many hospitals with acute care settings.

Acknowledgments

We thank Motoe Yumioka, Satomi Unno, Kazuyo Okushio, Sumiko Sano, Hideko Mochizuki, and Kazue Sano for their guidance in conducting this study. We thank Ms Etsuko Nagahashi and Mr Hiroki Ueda of the Information Support Office of Shizuoka General Hospital for assistance with data acquisition. We thank Katherine Thieltges and Analisa Avila, MPH, ELS, from Edanz for editing a draft of the manuscript. This study was supported by a grant-in-aid for Medical Research Promotion from the Shizuoka Prefectural Hospital Organization and a research instruction budget from the Shizuoka Graduate University of Public Health.

Data Availability

The datasets generated and analyzed during this study are not publicly available since the authors failed to obtain informed consent regarding the disclosure of raw patient data at the start of the study, which means the data in the manuscript, supporting information files, or public repositories cannot be freely accessed. However, the datasets are available from the Clinical Trial Management Office at Shizuoka General Hospital, located at 4-27-1 Kita-ando, Aoi-ku, Shizuoka, Japan 420-8527 (chiken-sougou@shizuoka-pho.jp), on reasonable request.

Authors' Contributions

CS and EN contributed to study concept and design and data analysis; CS and KH contributed to data acquisition; CS, EN, HS, NEK, MT, and KH did data interpretation; and CS contributed to first draft of the manuscript. All authors read drafts of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figure.
[DOCX File, 151 KB - humanfactors_v12i1e58073_app1.docx]

References

1. Ministry of Health, Labour and Welfare. White paper on aging society [Article in Japanese]. Cabinet Office. 2023. URL: https://www8.cao.go.jp/kourei/whitepaper/w-2023/zenbun/pdf/1s1s_01.pdf [accessed 2024-12-13]
2. Status of inpatients by category of basic charge for hospitalization in general ward [Article in Japanese]. Ministry of Health, Labour and Welfare. 2017. URL: https://www.mhlw.go.jp/file/05-Shingikai-12404000-Hokenkyoku-Iryouka/0000155341.pdf [accessed 2024-12-13]
3. Dykes PC, Curtin-Bowen M, Lipsitz S, et al. Cost of inpatient falls and cost-benefit analysis of implementation of an evidence-based fall prevention program. JAMA Health Forum 2023 Jan 6;4(1):e225125. [doi: 10.1001/jamahealthforum.2022.5125] [Medline: 36662505]
4. Aihara H, Tago M, Oishi T, Katsuki NE, Yamashita SI. Visual impairment, partially dependent ADL and extremely old age could be predictors for severe fall injuries in acute care settings. Int J Gerontol 2018 Sep;12(3):175-179. [doi: 10.1016/j.ijge.2018.02.014]
5. Panel on Prevention of Falls in Older Persons, American Geriatrics Society and British Geriatrics Society. Summary of the updated American Geriatrics Society/British Geriatrics Society clinical practice guideline for prevention of falls in older persons. J Am Geriatr Soc 2011 Jan;59(1):148-157. [doi: 10.1111/j.1532-5415.2010.03234.x]
6. Cuevas-Trisan R. Balance problems and fall risks in the elderly. Clin Geriatr Med 2019 May;35(2):173-183. [doi: 10.1016/j.cger.2019.01.008] [Medline: 30929881]

7. Canning CG, Sherrington C, Lord SR, et al. Exercise for falls prevention in Parkinson disease: a randomized controlled trial. Neurology (ECronicon) 2015 Jan 20;84(3):304-312. [doi: 10.1212/WNL.0000000000001155] [Medline: 25552576]
8. Hill AM, Francis-Coad J, Haines TP, et al. "My independent streak may get in the way": how older adults respond to falls prevention education in hospital. BMJ Open 2016 Jul 26;6(7):e012363. [doi: 10.1136/bmjopen-2016-012363] [Medline: 27466244]
9. Heng H, Kiegaldie D, Slade SC, et al. Healthcare professional perspectives on barriers and enablers to falls prevention education: a qualitative study. PLoS One 2022;17(4):e0266797. [doi: 10.1371/journal.pone.0266797] [Medline: 35476840]
10. Morris ME, Webster K, Jones C, et al. Interventions to reduce falls in hospitals: a systematic review and meta-analysis. Age Ageing 2022 May 1;51(5):afac077. [doi: 10.1093/ageing/afac077] [Medline: 35524748]
11. Shaw L, Kiegaldie D, Morris ME. Educating health professionals to implement evidence-based falls screening in hospitals. Nurse Educ Today 2021 Jun;101:104874. [doi: 10.1016/j.nedt.2021.104874] [Medline: 33774527]
12. Heng H, Jazayeri D, Shaw L, Kiegaldie D, Hill AM, Morris ME. Hospital falls prevention with patient education: a scoping review. BMC Geriatr 2020 Apr 15;20(1):140. [doi: 10.1186/s12877-020-01515-w] [Medline: 32293298]
13. Kawazoe Y, Shimamoto K, Shibata D, Shinohara E, Kawaguchi H, Yamamoto T. Impact of a clinical text-based fall prediction model on preventing extended hospital stays for elderly inpatients: model development and performance evaluation. JMIR Med Inform 2022 Jul 27;10(7):e37913. [doi: 10.2196/37913] [Medline: 35896017]
14. Cho I, Boo EH, Chung E, Bates DW, Dykes P. Novel approach to inpatient fall risk prediction and its cross-site validation using time-variant data. J Med Internet Res 2019 Feb 19;21(2):e11505. [doi: 10.2196/11505] [Medline: 30777849]
15. Cho I, Jin IS, Park H, Dykes PC. Clinical impact of an analytic tool for predicting the fall risk in inpatients: controlled interrupted time series. JMIR Med Inform 2021 Nov 25;9(11):e26456. [doi: 10.2196/26456] [Medline: 34626168]
16. Jung H, Yoo S, Kim S, et al. Patient-level fall risk prediction using the observational medical outcomes partnership's common data model: pilot feasibility study. JMIR Med Inform 2022 Mar 11;10(3):e35104. [doi: 10.2196/35104] [Medline: 35275076]
17. Oliver D, Britton M, Seed P, Martin FC, Hopper AH. Development and evaluation of evidence based risk assessment tool (STRATIFY) to predict which elderly inpatients will fall: case-control and cohort studies. BMJ 1997 Oct 25;315(7115):1049-1053. [doi: 10.1136/bmj.315.7115.1049] [Medline: 9366729]
18. Morse JM, Morse RM, Tylko SJ. Development of a scale to identify the fall-prone patient. Can J Aging 1989;8(4):366-377. [doi: 10.1017/S0714980800008576]
19. Beasley B, Patatanian E. Development and implementation of a pharmacy fall prevention program. Hosp Pharm 2009 Dec;44(12):1095-1102. [doi: 10.1310/hpj4412-1095]
20. Tago M, Katsuki NE, Oda Y, Nakatani E, Sugioka T, Yamashita SI. New predictive models for falls among inpatients using public ADL scale in Japan: a retrospective observational study of 7,858 patients in acute care setting. PLoS One 2020;15(7):e0236130. [doi: 10.1371/journal.pone.0236130] [Medline: 32673366]
21. Hirata R, Katsuki NE, Yaita S, et al. Validation of the Saga Fall Injury Risk Model. Int J Med Sci 2024;21(8):1378-1384. [doi: 10.7150/ijms.92837] [Medline: 38903917]
22. Yaita S, Tago M, Katsuki NE, et al. A simple and accurate model for predicting fall injuries in hospitalized patients: insights from a retrospective observational study in Japan. Med Sci Monit 2023 Aug 14;29:e941252. [doi: 10.12659/MSM.941252] [Medline: 37574766]
23. Nakatani H, Nakao M, Uchiyama H, Toyoshiba H, Ochiai C. Predicting inpatient falls using natural language processing of nursing records obtained from Japanese electronic medical records: case-control study. JMIR Med Inform 2020 Apr 22;8(4):e16970. [doi: 10.2196/16970] [Medline: 32319959]
24. Soyano A, Suzuki M, Harada A, et al. JSFP Research Report〉 fall risk assessment tool usage and issues: results of a fall risk assessment tool usage and issues: results of a. Jpn J Fall Prev 2018;5:41-49. [doi: 10.11335/tentouyobou.5.1_41]
25. Higaonna M. The predictive validity of a modified Japanese Nursing Association fall risk assessment tool: a retrospective cohort study. Int J Nurs Stud 2015 Sep;52(9):1484-1494. [doi: 10.1016/j.ijnurstu.2015.05.009] [Medline: 26138481]
26. Okuma Y. Freezing of gait and falls in Parkinson's disease. J Parkinsons Dis 2014;4(2):255-260. [doi: 10.3233/JPD-130282] [Medline: 24577502]
27. Hodgson G, Pace A, Carfagnini Q, et al. Risky business: factors that increase risk of falls among older adult in-patients. Gerontol Geriatr Med 2023;9:23337214231189930. [doi: 10.1177/23337214231189930] [Medline: 37533770]
28. Hendrich AL, Bender PS, Nyhuis A. Validation of the Hendrich II Fall Risk Model: a large concurrent case/control study of hospitalized patients. Appl Nurs Res 2003 Feb;16(1):9-21. [doi: 10.1053/apnr.2003.YAPNR2] [Medline: 12624858]
29. Obayashi K, Araki T, Nakamura K, et al. Risk of falling and hypnotic drugs: retrospective study of inpatients. Drugs R D 2013 Jun;13(2):159-164. [doi: 10.1007/s40268-013-0019-3] [Medline: 23760758]
30. Noh HM, Song HJ, Park YS, Han J, Roh YK. Fall predictors beyond fall risk assessment tool items for acute hospitalized older adults: a matched case-control study. Sci Rep 2021 Jan 15;11(1):1503. [doi: 10.1038/s41598-021-81034-9] [Medline: 33452370]
31. Nakamura K, Kitamura K, Watanabe Y, et al. Body mass index and risk of recurrent falls in community-dwelling Japanese aged 40-74 years: the Murakami cohort study. Geriatr Gerontol Int 2021 Jun;21(6):498-505. [doi: 10.1111/ggi.14167] [Medline: 33881216]

32. Ylitalo KR, Karvonen-Gutierrez CA. Body mass index, falls, and injurious falls among U.S. adults: findings from the 2014 Behavioral Risk Factor Surveillance System. Prev Med 2016 Oct;91:217-223. [doi: 10.1016/j.ypmed.2016.08.044] [Medline: 27575319]
33. Yi SW, Kim YM, Won YJ, Kim SK, Kim SH. Association between body mass index and the risk of falls: a nationwide population-based study. Osteoporos Int 2021 Jun;32(6):1071-1078. [doi: 10.1007/s00198-020-05725-1] [Medline: 33411009]
34. Trevisan C, Crippa A, Ek S, et al. Nutritional status, body mass index, and the risk of falls in community-dwelling older adults: a systematic review and meta-analysis. J Am Med Dir Assoc 2019 May;20(5):569-582. [doi: 10.1016/j.jamda.2018.10.027] [Medline: 30554987]
35. Ogliari G, Ryg J, Andersen-Ranberg K, Scheel-Hincke LL, Masud T. Association between body mass index and falls in community-dwelling men and women: a prospective, multinational study in the Survey of Health, Ageing and Retirement in Europe (SHARE). Eur Geriatr Med 2021 Aug;12(4):837-849. [doi: 10.1007/s41999-021-00485-5] [Medline: 33817765]
36. Tago M, Katsuki NE, Yaita S, et al. High inter-rater reliability of Japanese bedriddenness ranks and cognitive function scores: a hospital-based prospective observational study. BMC Geriatr 2021 Mar 9;21(1):168. [doi: 10.1186/s12877-021-02108-x] [Medline: 33750305]
37. Ozawa MN, Nakayama S. Long-term care insurance in Japan. J Aging Soc Policy 2005;17(3):61-84. [doi: 10.1300/J031v17n03_04] [Medline: 16219594]
38. Council for Medical Safety Management of National University Hospitals influence classification [Article in Japanese]. Japan National University Hospital Alliance on Patient Safety. URL: https://square.umin.ac.jp/anzenhc/information/pdf/23_business_text2022.pdf [accessed 2024-12-13]
39. Kanda Y. Investigation of the freely available easy-to-use software "EZR" for medical statistics. Bone Marrow Transplant 2013 Mar;48(3):452-458. [doi: 10.1038/bmt.2012.244] [Medline: 23208313]
40. Hirata R, Tago M, Katsuki NE, et al. History of falls and bedriddenness ranks are useful predictive factors for in-hospital falls: a single-center retrospective observational study using the Saga Fall Risk Model. Int J Gen Med 2022;15:8121-8131. [doi: 10.2147/IJGM.S385168] [Medline: 36389017]
41. Masumoto T, Yamada Y, Yamada M, et al. Fall risk factors and sex differences among community-dwelling elderly individuals in Japan. A Kameoka study. Nihon Koshu Eisei Zasshi 2015;62(8):390-401. [doi: 10.11236/jph.62.8_390] [Medline: 26511610]
42. Bally ELS, Ye L, van Grieken A, et al. Factors associated with falls among hospitalized and community-dwelling older adults: the APPCARE study. Front Public Health 2023;11:1180914. [doi: 10.3389/fpubh.2023.1180914] [Medline: 37457268]
43. Tsuboi A, Murakami T, Kurumadani H, et al. Relationship between criteria for evaluating the degree of independence of disabled elderly persons in performing activities of daily living. J Jpn Occup Ther Assoc 2002;21:255-262. [doi: 10.11477/mf.6003900624]
44. Haines TP, Hill AM, Hill KD, et al. Patient education to prevent falls among older hospital inpatients: a randomized controlled trial. Arch Intern Med 2011 Mar 28;171(6):516-524. [doi: 10.1001/archinternmed.2010.444] [Medline: 21098343]
45. Montero-Odasso M, van der Velde N, Martin FC, et al. World guidelines for falls prevention and management for older adults: a global initiative. Age Ageing 2022 Sep 2;51(9):afac205. [doi: 10.1093/ageing/afac205] [Medline: 36178003]
46. Jellett J, Williams C, Clayton D, Plummer V, Haines T. Falls risk score removal does not impact inpatient falls: a stepped-wedge, cluster-randomised trial. J Clin Nurs 2020 Dec;29(23-24):4505-4513. [doi: 10.1111/jocn.15471] [Medline: 32945020]
47. Morris ME, Haines T, Hill AM, et al. Divesting from a scored hospital Fall Risk Assessment Tool (FRAT): a cluster randomized non-inferiority trial. J Am Geriatr Soc 2021 Sep;69(9):2598-2604. [doi: 10.1111/jgs.17125] [Medline: 33834490]
48. Swift CG, Iliffe S. Assessment and prevention of falls in older people—concise guidance. Clin Med (Northfield) 2014 Dec;14(6):658-662. [doi: 10.7861/clinmedicine.14-6-658]
49. Wong CL, Holroyd-Leduc J, Simel DL, Straus SE. Does this patient have delirium?: value of bedside instruments. JAMA 2010 Aug 18;304(7):779-786. [doi: 10.1001/jama.2010.1182] [Medline: 20716741]
50. Inouye SK, Westendorp RGJ, Saczynski JS. Delirium in elderly people. Lancet 2014 Mar;383(9920):911-922. [doi: 10.1016/S0140-6736(13)60688-1]
51. Herzig SJ, Rothberg MB, Moss CR, et al. Risk of in-hospital falls among medications commonly used for insomnia in hospitalized patients. Sleep 2021 Sep 13;44(9):zsab064. [doi: 10.1093/sleep/zsab064] [Medline: 33710329]
52. Kojima T, Akishita M, Nakamura T, et al. Association of polypharmacy with fall risk among geriatric outpatients. Geriatr Gerontol Int 2011 Oct;11(4):438-444. [doi: 10.1111/j.1447-0594.2011.00703.x] [Medline: 21545384]
53. Zaninotto P, Huang YT, Di Gessa G, Abell J, Lassale C, Steptoe A. Polypharmacy is a risk factor for hospital admission due to a fall: evidence from the English Longitudinal Study of Ageing. BMC Public Health 2020 Nov 26;20(1):1804. [doi: 10.1186/s12889-020-09920-x] [Medline: 33243195]
54. Dhalwani NN, Fahami R, Sathanapally H, Seidu S, Davies MJ, Khunti K. Association between polypharmacy and falls in older adults: a longitudinal study from England. BMJ Open 2017 Oct 16;7(10):e016358. [doi: 10.1136/bmjopen-2017-016358] [Medline: 29042378]

Abbreviations

ADL: activities of daily living
BR: bedriddenness rank
EHR: electronic health record
HR: hazard ratio

Edited by P Santana-Mancilla; submitted 13.04.24; peer-reviewed by S Haag, C Udeh, L Lehtonen; revised version received 24.09.24; accepted 30.09.24; published 13.01.25.

Please cite as:
Saito C, Nakatani E, Sasaki H, E Katsuki N, Tago M, Harada K
Predictive Factors and the Predictive Scoring System for Falls in Acute Care Inpatients: Retrospective Cohort Study
JMIR Hum Factors 2025;12:e58073
URL: https://humanfactors.jmir.org/2025/1/e58073
doi:10.2196/58073

Original Paper

An Explainable AI Application (AF'fective) to Support Monitoring of Patients With Atrial Fibrillation After Catheter Ablation: Qualitative Focus Group, Design Session, and Interview Study

Wan Jou She[1], PhD; Panote Siriaraya[1], PhD; Hibiki Iwakoshi[2,3], MD, PhD; Noriaki Kuwahara[1,4], PhD; Keitaro Senoo[2,3], MD, PhD

[1]Faculty of Information and Human Sciences, Kyoto Institute of Technology, Kyoto, Japan

[2]Department of Cardiovascular Medicine, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan

[3]Department of Cardiac Arrhythmia Research and Innovation, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan

[4]Department of Advanced Fibro-Science, Kyoto Institute of Technology, Kyoto, Japan

Corresponding Author:
Keitaro Senoo, MD, PhD
Department of Cardiac Arrhythmia Research and Innovation
Graduate School of Medical Science
Kyoto Prefectural University of Medicine
465 Kajii-cho Hirokoji, Kawaramachi-dori
Kamigyo-ku Kyoto
Kyoto, 602-8566
Japan
Phone: 81 75 251 5511
Fax: 81 75 251 5514
Email: k-senoo@koto.kpu-m.ac.jp

Abstract

Background: The opaque nature of artificial intelligence (AI) algorithms has led to distrust in medical contexts, particularly in the treatment and monitoring of atrial fibrillation. Although previous studies in explainable AI have demonstrated potential to address this issue, they often focus solely on electrocardiography graphs and lack real-world field insights.

Objective: We addressed this gap by incorporating standardized clinical interpretation of electrocardiography graphs into the system and collaborating with cardiologists to co-design and evaluate this approach using real-world patient cases and data.

Methods: We conducted a 3-stage iterative design process with 23 cardiologists to co-design, evaluate, and pilot an explainable AI application. In the first stage, we identified 4 physician personas and 7 explainability strategies, which were reviewed in the second stage. A total of 4 strategies were deemed highly effective and feasible for pilot deployment. On the basis of these strategies, we developed a progressive web application and tested it with cardiologists in the third stage.

Results: The final progressive web application prototype received above-average user experience evaluations and effectively motivated physicians to adopt it owing to its ease of use, reliable information, and explainable functionality. In addition, we gathered in-depth field insights from cardiologists who used the system in clinical contexts.

Conclusions: Our study identified effective explainability strategies, emphasized the importance of curating actionable features and setting accurate expectations, and suggested that many of these insights could apply to other disease care contexts, paving the way for future real-world clinical evaluations.

(JMIR Hum Factors 2025;12:e65923) doi:10.2196/65923

KEYWORDS

atrial fibrillation; explainable artificial intelligence; explainable AI; user-centered design; prevention; postablation monitoring

Introduction

Background

While the role of artificial intelligence (AI) in health care is growing, concerns about its reliability and transparency remain largely unaddressed [1]. The opaque nature of AI algorithms challenges trust, particularly in cardiology, where the lack of interpretability in deep learning models limits their adoption and utility [2,3]. This has led to increased interest in developing explainable AI models that offer accurate predictions along with clear explanations of the rationale behind the predictive results [4].

Our study explored the advantages and challenges of using AI models in cardiovascular treatment, focusing on designing an explainable AI system to aid posttreatment care for patients with atrial fibrillation (AF) after catheter ablation. There were >3 million new cases of AF, a common arrhythmic disorder [5], in 2017, with numbers expected to rise due to aging [6]. Although often undetected, AF is a significant risk factor for severe conditions such as stroke or heart failure [7] and can greatly impact patients' quality of life [5,8]. While catheter ablation is effective, AF can recur after ablation, requiring careful monitoring of risk factors [9].

Several machine learning models, from screening algorithms [10,11] to those predicting AF recurrence after ablation [12], have been developed, but integrating them into health care remains challenging. Physicians often require detailed justifications due to prediction uncertainties. While explainable AI algorithms [5,13] could assist, a more holistic approach is needed in posttreatment scenarios to effectively integrate additional information, which remains poorly understood in real-world care contexts (discussed in the next section).

To better integrate a machine learning model into the ongoing decision-making ecosystem and identify unmet clinical needs beyond performance optimization, we used a user-centered, iterative design approach to develop AF'fective, an explainable AI system for posttreatment monitoring of patients with AF. By conducting semistructured interviews, design sessions, and focus groups with 23 cardiologists, we aimed to identify key principles for the effective design and use of explainable AI in real-life cardiac treatment and management. In this section, we first review the development of explainable AI in relevant fields in the second subsection and summarize our first 2 study stages in the third subsection. We then detail and discuss our final-stage findings and conclude with reflections on the study's implications and the generalizability of our insights.

High Accuracy but (Still) Rejected: Machine Learning Models in Cardiology

Cardiology is one field of medicine that has seen extensive use of AI to aid in medical practice. Various machine learning models have been developed and used to assist in a wide range of tasks, from helping predict the prognosis or readmission after heart failure [14,15] to detecting various cardiovascular diseases from medical images [16]. In the area of arrhythmia, previous studies have focused particularly on two key areas: (1) prevention, which involves assisting in the early detection of various arrhythmic conditions; and (2) monitoring, which involves supporting the management of such conditions after treatment [17]. In the preventive area, researchers have shown how machine learning models could be developed and used to detect irregular conditions using electrocardiography (ECG) signals [18] or even through photoplethysmography sensors from commercially available smartwatches [19,20]. In most cases, such models were able to diagnose these irregular conditions with a high degree of accuracy (97%) [20]. However, the potential for a high rate of false positives when implemented through commercial smartwatches has triggered concerns among researchers, and as a result, they have cautioned against using such systems for population-wide screening and emphasized maintaining the 12-lead ECG as the gold standard [21]. Given the prevalence of AF, several of the developed models in this domain have chosen to focus specifically on this condition, with various deep learning ECG models being developed to help screen people for AF, stratify patients based on their risk level, and even predict the chances of the condition occurring in the future (see the study by Sehrawat et al [21] for a comprehensive review).

In the monitoring area, to support posttreatment care, previous studies have developed models capable of predicting patient mortality or echocardiographic response after procedures such as cardiac resynchronization therapy [22]. Other studies have even shown how the risk of recurrence for conditions such as AF could be predicted using patient demographics and 3D computed tomography images of the left atrium [23]. Similarly, several studies have demonstrated how the recurrence of AF or 30-day hospital readmission after catheter ablation can be predicted using various deep learning and non–deep learning strategies (eg, convolutional neural network [CNN] and Extreme Gradient Boosting) [12,13,24]. In other cases, deep learning algorithms have also been developed to detect an AF episode 4.5 minutes before onset, thus enabling prompt interventions to prevent their occurrence [25].

Overall, while the findings of these studies all appear to demonstrate significant potential for implementing predictive models to support AF treatment, we have encountered substantial pushback from clinical partners when attempting to implement such a model in clinical practice. Despite the promising performance of the model in the study by Nishimura et al [26] (area under the curve of 0.72 with 83% sensitivity and 58% specificity), there was still strong resistance from physicians regarding trusting the predicted outcomes. Such hesitation and resistance to adopting machine learning models among clinical staff have also been frequently identified in previous studies [1,27]. Given the delicate and often critical nature of treatment in this field, it is not surprising that the lack of interpretability undermines trust in these models, preventing clinicians from using them in medical treatment (see the study by Petch et al [2]). Moreover, the opaque process raises concerns regarding how mistakes could be rectified, not to mention that it may cause potentially disruptive emotions when discussing such results with patients. Therefore, at present, the high accuracy, accessibility, and (possibly) low cost of machine learning models often do not lead to their adoption in medical practice.

Explainable AI in AF

Recently, explainable AI has gained attention as a potential approach to address these issues of transparency and interpretability [28-30]. In the context of AF treatment, ECG data are generally used as a main data source for screening irregular rhythm and predicting the risk of AF recurrence, and as such, explainable AI techniques such as Shapley Additive Explanations (SHAP) have been used with the ECG data to identify features contributing to the predictive outcomes [31]. In previous studies, researchers have proposed an explainable deep learning model that not only detects AF but also describes the reasons behind their decisions and visualizes key regions within the ECG signal identified as important predictors using techniques such as Gradient-Weighted Class Activation Mapping (Grad-CAM) or a deep visual attention [29,32,33]. Building upon earlier models, researchers such as Raza et al [34] incorporated federated learning techniques alongside explainable AI to preserve patient ECG data privacy. In more recent studies, explainable models have even been created that are capable of identifying patients at high risk after catheter ablation and highlighting the features (eg, type of AF, age, and left atrial diameter) that the model used to decide on the risk level for each patient [30]. The use of "explainable" models in this manner has been argued to have the added benefit of allowing clinicians to better understand the relationship between each contributing factor and the predicted outcome of an individual case, helping clinicians justify their decisions and treatments more effectively [35].

While explainable AI techniques are believed to enhance the interpretability of algorithms, incorporating them to truly improve patient outcomes in real-life medical contexts remains challenging [36]. Understanding of the strengths and limitations of these approaches in different use cases and how human and AI-based diagnoses can complement each other is still lacking [36]. Deploying these models effectively requires a thorough examination of the patient journey and adapting the explainable strategy accordingly. Recent studies [37,38] have begun to address these issues by examining the use of explainable AI in more realistic settings with various stakeholders to ensure safe adoption in clinical practice.

Overall, although the integration of explainable AI into medical contexts holds great promise, it requires a concerted effort to address technical, ethical, and practical challenges. Insights from near-live and real-life case studies can help develop best practices and trustworthy methods for implementing such systems in clinical settings, thereby enabling users and professionals to fully benefit from AI technology.

Early Study Stages and Findings in a Nutshell

In this section, we summarize and report the in-depth insights from the 2 early stages of interviews (partially reported as a poster in the study by She et al [39]) and group design sessions to conceptualize and pilot-test our design strategies and prototype. We chose to include some key findings from these early stages to enhance readers' understanding of our design decisions in developing the final prototype for evaluation.

Stage 1

Overview

In the first stage, we conducted semistructured interviews with physicians to gain insights into their experiences and best practices for posttreatment care of patients with AF. Physicians were asked to share their positive and negative experiences, communication strategies, and effective procedures for monitoring posttreatment patients with AF. To sensitize and prepare participants, we provided a 1-week workbook with reflection exercises sent 1 to 2 weeks before the interview. Each interview lasted approximately 45 minutes and was fully recorded with the consent of the physicians.

Participants

We invited 8 physicians from different specialties, such as emergency medicine, internal medicine, and cardiology, to join us for interviews at Kyoto Prefectural University of Medicine in Japan. We made sure to include a mix of experience levels and backgrounds in cardiology to obtain insights that could be more broadly generalized to other medical contexts.

Measurements and Study Apparatus

A week before the interview, each physician received a toolkit to reflect on their clinical practice. The toolkit included a workbook with 7 daily assignments on their experiences and patient interactions as well as stickers featuring emotional words and contextual images to help express their thoughts and feelings.

Study Process

We asked physicians to reflect on their process for diagnosing patients and explaining treatment plans, especially in difficult situations such as delivering bad news. They were also asked to share strategies for handling uncooperative patients or those who rejected treatment. Physicians were encouraged to use the 2 sticker sets to respond to workbook questions or use them as interview resources. During the interview, physicians were guided through the workbook to discuss their specific approaches, experiences, motivations, stressors, and methods for negotiating treatment plans.

Analysis

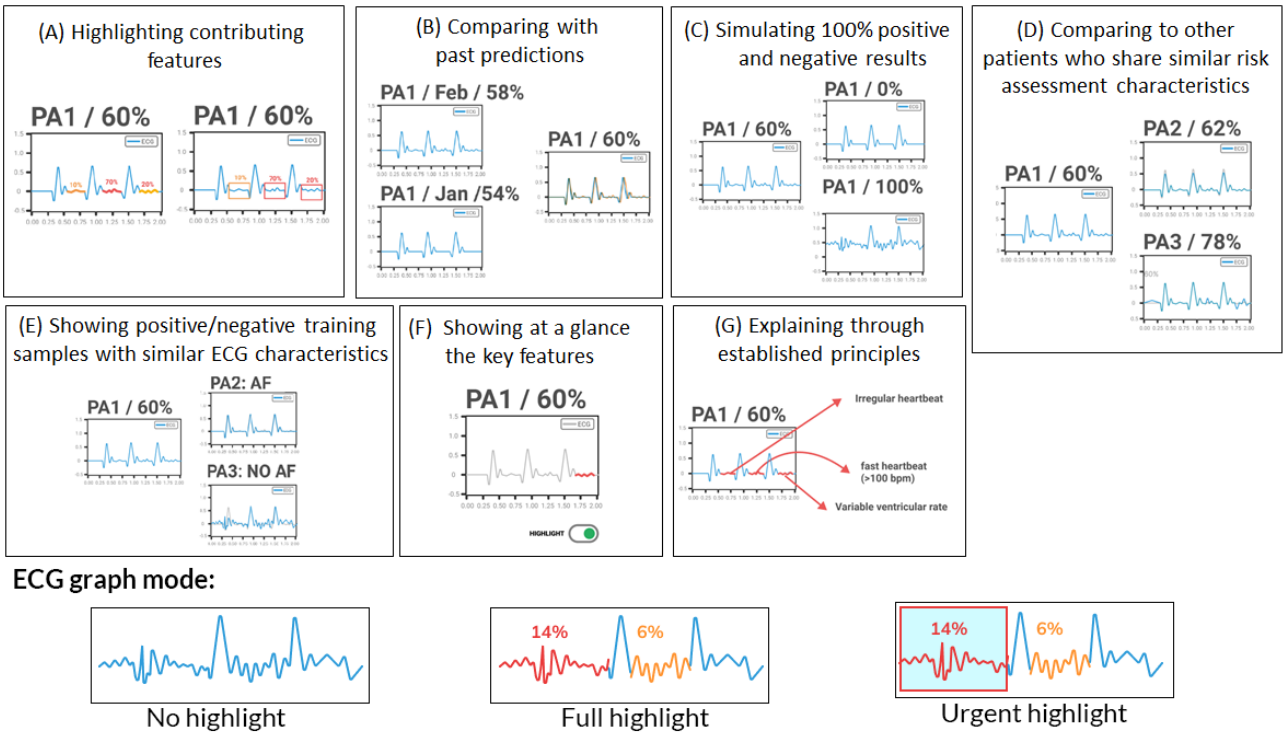
In total, 2 researchers from the human-computer interaction (HCI) field and a cardiologist analyzed the interview data, with all recordings fully transcribed. The HCI researchers developed design strategies through thematic analysis, whereas the cardiologist reviewed and selected strategies for further evaluation in stage 2.

Results

Our initial interviews with 8 physicians produced 2 key outputs: 4 physician personas (*teamworker*, *minimum effort professional*, *the veteran*, and *self-actualizer*) and 7 design strategies for explainable prediction. Each persona included pseudodemographics and progress bars depicting their personalities across 4 parameters: *passionate/realistic*, *rookie/veteran*, *work-oriented/life-oriented*, and *individualism/collectivism*.

The research team discovered that participants had common concerns about self-efficacy in patient treatment, feeling validated when treatments succeeded but disempowered when seen as incompetent. This insight led to the development of 7 design strategies for explainable prediction. Participants emphasized the need to justify decisions by clearly explaining issues such as ECG results to patients. We reviewed and adapted explainable AI strategies from previous studies [40] for cardiology. The seven strategies were (1) highlighting key ECG features (strategy A), (2) comparing with historical predictions (strategy B), (3) simulating extreme ECG outcomes (strategy C), (4) comparing with similar patients (strategy D), (5) showing similar training examples (strategy E), (6) providing an at-a-glance mode (strategy F), and (7) explaining through established ECG principles (strategy G; Figure 1 [39]).

Figure 1. The 7 explainable artificial intelligence strategies (A-G) that were examined in the initial 2 stages of the study, expanded from the previous publication [39]. AF: atrial fibrillation; bpm: beats per minute; ECG: electrocardiography; PA1: patient 1; PA2: patient 2; PA3: patient 3.



Stage 2

Overview

In the second stage, we evaluated the feasibility of the 7 design strategies from the first stage, focusing on those that could be implemented using explainable AI. Physicians first described their usual process for assessing recurrence risk and discussed the potential role of predictive models and explainable AI. We then introduced the explainable AI strategies and gathered detailed feedback on which would most effectively clarify predicted outcomes for other physicians and patients.

Study Process and Findings to Guide the Prototype Development

We conducted 2 group design sessions with 8 cardiologists to critically review and assess our design strategies. Initially, 5 cardiologists were asked to evaluate the advantages and disadvantages of each strategy, describe scenarios in which they would be useful, and reflect on their concerns regarding implementing such explainable features.

From the 7 proposed strategies, we selected 4 (strategies A, B, D, and G) for further development into a web-based prototype using a high-fidelity explainable CNN model with real patient data. In total, 3 cardiologists reviewed the prototype using the concurrent think-aloud approach [41] and provided feedback. They also discussed concerns about using AI during posttreatment care, specifically on whether the explainable features enhanced their trust in the AI system and how the system might add or reduce value in their practice.

Participants

A total of 5 cardiologists with experience treating AF were recruited from Kyoto Prefectural University of Medicine to assess the explainable AI strategies. In addition, 3 cardiologists from the same institution were brought on to evaluate a web-based prototype.

Results and Study Apparatus Development

The 4 effective design strategies identified were highlighting contributing features (strategy A), comparing with past predictors (strategy B), comparing with other patients (strategy D), and explaining through established principles (strategy G). Participants' choices were guided by 3 principles: discernibility (highlighting relevant noteworthy parts), comparability (intra- and interpersonal comparisons), and evidence-based approaches (derived from diagnostic examples or clinical literature).

Strategies to avoid included non–evidence-based methods and those showing only partial graphs.

In general, physicians were highly receptive to the AF'fective prototype, particularly its explainable features, which they found useful for evaluating treatment effectiveness in posttreatment monitoring. On the basis of their feedback, we advanced to web-based prototyping.

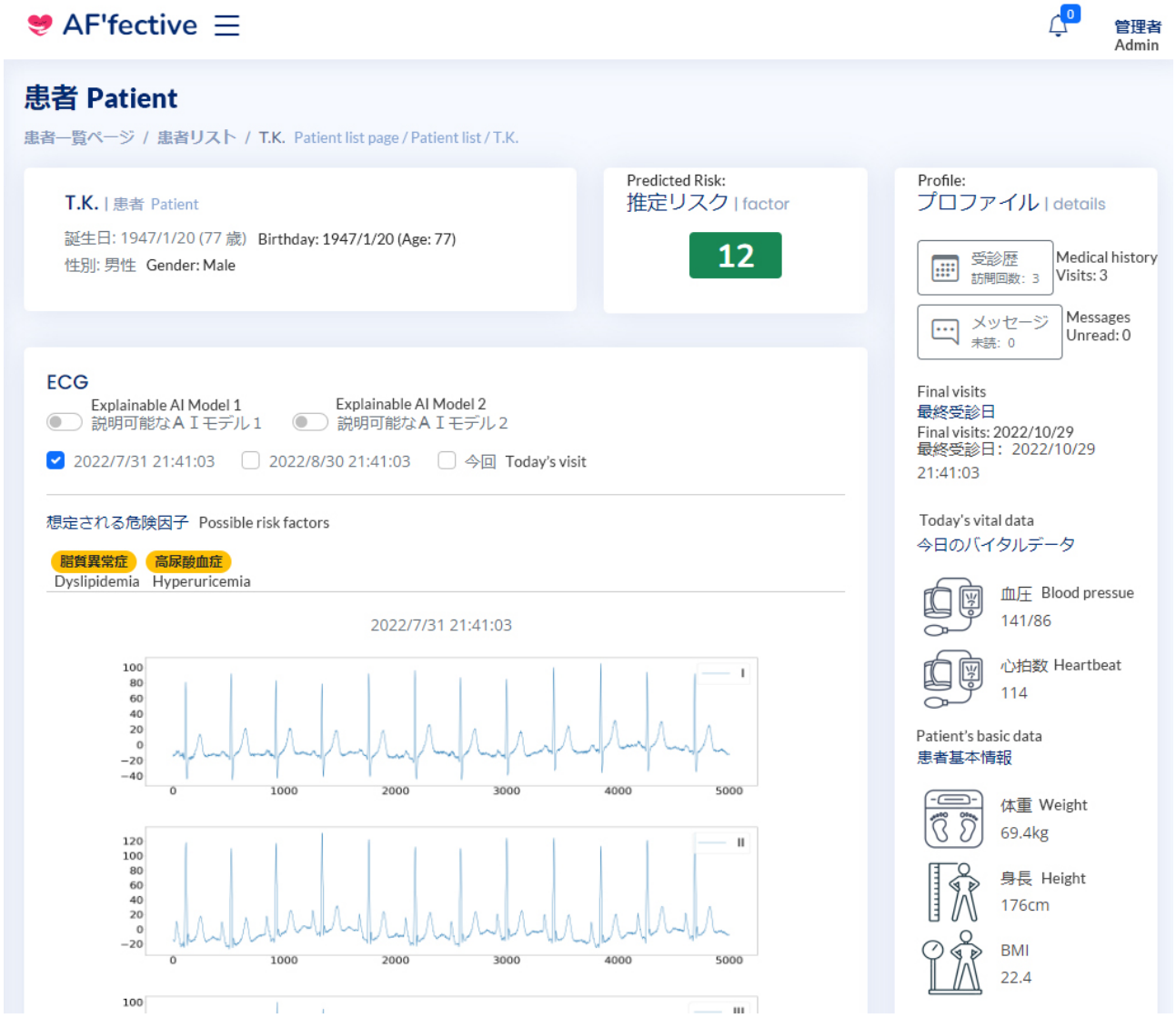
The web-based prototype was developed using Figma (Figma, Inc), a popular tool among user experience (UX) designers for creating high-fidelity and web-based prototypes at minimal cost during the development process. Our prototype comprised 2 main screens: a patient overview screen and an individual health record dashboard (to avoid repetition of similar images, we share the screenshots from the final system in this paper instead; Figures 2 and 3).

The overview screen displays patients' demographic details and predicted recurrence risks, allowing cardiologists to sort by various demographic and risk factors. The health record dashboard displays (comparable and evidence-based) health information gathered from a patient's previous visits. Our predictive model will automatically extract and highlight keywords or risk factors based on the patient's visits and health data. In addition, the model differentiates the importance of the risk factors and offers 3 tiers of ECG result presentation based on the level of urgency for physicians: no highlight, full highlight, and critical highlight only.

In critical highlight mode, only urgent health information is shown, whereas the full highlight mode displays all details, including less urgent AI highlights. Validated risk factors are listed as diagnostic keywords in the right panel, helping physicians quickly identify the main contributors to the predictive outcome.

Figure 2. The patient overview screen.

Figure 3. Screenshot of the patient page, including extra risk factors. ECG: electrocardiography.



Methods

Overview

Our research followed a 3-phase approach collaborating with stakeholders familiar with AF management. In the first phase, we conducted semistructured interviews with 8 cardiologists to explore potential explainable AI strategies for postablation care. From these insights, we developed 7 design concepts, which were then critically assessed and refined by another 8 cardiologists in the second phase. This process led to the selection of 4 explainable strategies. In this section (the final phase of the study), we built a functional prototype incorporating real patient data and invited a third group of 7 cardiologists to evaluate its effectiveness and feasibility in clinical settings.

Ethical Considerations

This study received ethics approval from the relevant institutional review boards (IRBs) in Japan, including the Kyoto Institute of Technology IRB (2021-08) and the Kyoto Prefectural University of Medicine IRB (ERB-C-2014). Informed consent was obtained from all participating physicians, who volunteered to take part in the study without receiving any compensation. To ensure the privacy and confidentiality of patient health data, all personally identifiable information, including patients' names, was removed and replaced with unique codes. This anonymization process ensured that no data could be traced back to individual patients.

Stage 3

The web-based prototype from the previous stage was developed into a fully functional progressive web application (PWA) prototype, which was then tested and evaluated by cardiologists.

Patient Data

A total of 11 postablation patients consented to provide data for the prototype. Protected health information, such as names and birthdays, was masked or replaced due to IRB requests, but other health data (eg, ECG, body measurements, blood pressure, and visit history) were used to closely resemble the original patient health data for a realistic prototype.

Participants

In total, 7 cardiologists involved in the treatment of AF were recruited from the Kyoto Prefectural University of Medicine in

Japan. All of them had experience diagnosing patients with AF and performing catheter ablation.

Measurements and Study Apparatus

In total, 3 widely adapted UX-related questionnaires were implemented in our study: the User Experience Questionnaire (UEQ), technology acceptance model (TAM), and Mobile App Rating Scale (MARS). The rationales for including these questionnaires are explained in the following sections.

UEQ Inclusion Rationale

The UEQ, developed by Laugwitz et al [42] in 2008, consists of 26 items across 6 dimensions: attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty. The scale ranges from −3 (*most negative*) to +3 (*most positive*). Scores closer to 0 are considered "neutral." A data analysis tool is provided for interpreting the results, with benchmarks in 5 tiers: excellent, good, above average, below average, and bad [43]. Our goal during prototyping was to achieve above-average ratings. The English version was translated into Japanese.

TAM Inclusion Rationale

Previous research suggests that a user's intention to use is the primary predictor of actual system use [44]. To ensure that our PWA was both useful and usable in a medical context, we implemented the TAM to assess user acceptance and intention to use our explainable AI system. Originally developed by Davis et al [45], the TAM is a key model in new technology development. We adapted it for our PWA evaluation focusing on 3 dimensions: efficiency (perceived usefulness), ease of use (perceived ease of use), and intention to use (behavioral intention to use). For the TAM, a 7-point Likert scale is used, ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). A score of 4 represents neutrality.

MARS Inclusion Rationale

Despite its name, the MARS is widely used to assess the quality of health applications across various types of digital platforms. It evaluates 4 dimensions: engagement, functionality, esthetics, and information quality [46,47]. We selected the MARS to assess the information quality of our PWA using the Japanese translation by Yamamoto et al [48]. The MARS uses a 5-point Likert scale ranging from 1 (*inadequate*) to 5 (*excellent*). Scores of >3 are considered above average.

AF'fective: The Web-Based PWA

A fully web-based PWA was developed using the Ionic framework (Drifty) for the front end and Python (Python Software Foundation) for the back end. Access was strictly controlled through an authentication system to ensure data security. The PWA included 2 main pages that were developed based on our explainable strategies: a patient overview screen (corresponding to strategy D) and an individual health record dashboard (corresponding to strategies A, B, and G). One noteworthy change was made to our predictive model, which also influenced our explainable features. On the basis of physicians' feedback, we omitted direct highlights on the raw ECG graph in stage 2 (using Grad-CAM; see the demonstration in Figure 1 [39]) and, instead, used the 7-item features derived from the standardized clinical interpretation of ECG graphs (corresponding to strategies A and G in particular; see the demonstration in Figures 4 and 5). In particular, we replaced the explainable AI model used in previous stages with a SHAP model [49], which was constructed to highlight the extent to which different key clinical features—commonly used as established interpretation principles in AF prediction for ECG data (eg, maximum P wave duration and augmented vector right [aVR]/first precordial lead [V1])—contributed to the decision-making process for strategy G. Therefore, the predictive model was updated from a CNN model (image based) used in the previous 2 stages to a Cox regression model (feature based) [26] to accommodate the introduction of 7 standard features for prediction (and, later on, SHAP for explaining). While physicians were not shown the Cox regression model's performance to avoid biasing them, it achieved strong evaluations, with an area under the curve of 0.72, sensitivity of 83%, and specificity of 58%. The model was trained on data from 502 patients, with an average follow-up of 6.2 (SD 3.5) years. A total of 13.1% (66/502) of the patients developed new-onset AF [26].

These standardized clinical explainable features are included in the following list. These features were derived from the interpretation standards that cardiologists commonly apply when reading an ECG graph and assessing a patient's risk level. As indicated in our previous studies, while directly highlighting sections of the ECG graph that contributed to the predictive model's output can explain the model's prediction, it does not aid physicians in interpreting the graph. To enhance the interpretability and physician-friendly explanation, we asked cardiologists to standardize their interpretation of the ECG graph into higher-level features that are more straightforward and easier to understand for their fellow cardiologists. Our previous work [26] provides details on training a predictive model based on these new features:

1. Max P wave duration >125 ms: whether the maximum P wave duration is >125 ms
2. aVR/V1<1: whether the amplitude ratio of the P wave is <1
3. Amplitude V1≥10: whether the P wave amplitude in V1 is >0.1 mV
4. Amplitude aVR<4=0: whether the P wave amplitude in the lead aVR is <0.04 mV
5. PAC on ECG=1.0: whether there was one or more supraventricular ectopies during recording (removed in the latest model for simplicity from the work by Nishimura et al [26])
6. RV5SV1≥2.2=0: whether the amplitude (height) of the R wave in the fifth precordial lead plus the amplitude (depth) of the S wave in V1 is >2.2 mV
7. PR≥185=0: whether the interval between the P and R waves is >185 ms

The individual health record dashboard (as shown in Figure 3) also allowed physicians to check past visit ECG graphs and explainable features (corresponding to strategy B). The physicians can obtain an overview of the basic patient data in the top left corner and then proceed to the raw ECG graph. In the main graph section, they can choose between 2 types of explainable models (see Figures 4 and 5 for the corresponding

graphs) and add data from the patient's history for comparison. The right panel provides additional patient correspondence data and physical measurements such as weight and blood pressure. The standardized features made it easier to compare patients when physicians navigated the patient overview board (corresponding to strategy G).

We also developed 2 interfaces to summarize the explainable features. The first interface (Figure 4) was developed using SHAP, a popular tool for explaining the outcomes of machine learning models [49]. It features a single bar chart that shows the combined influence of risk and protective factors. The second interface uses a pie chart to show the relative importance of the 7 items and their risk or protective contributions to the predictive outcome (Figure 5). Cardiologists were advised to access it using a computer, although the site is also available on mobile devices, to simulate their normal use context.

Figure 4. Screenshot of the explainable items and how they were displayed in the system—explainable model 1. ECG: electrocardiography.

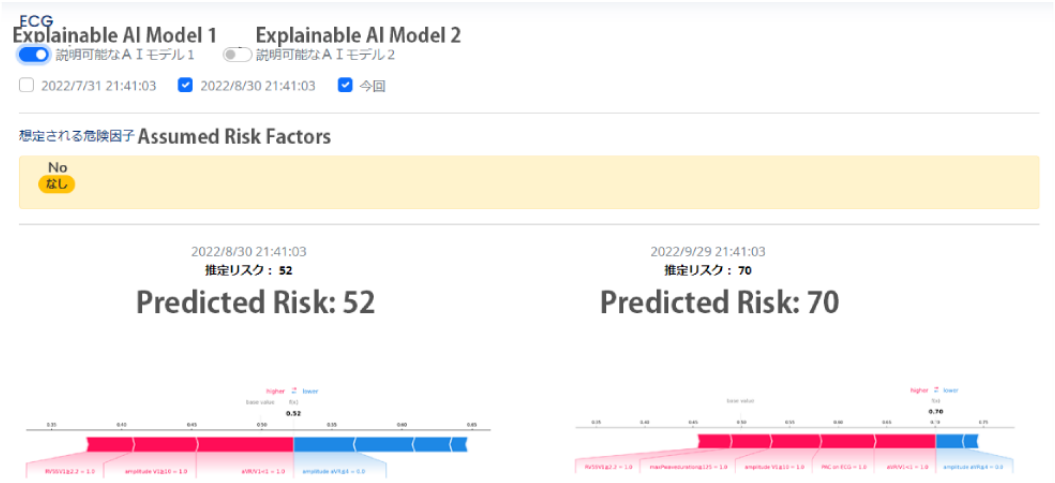
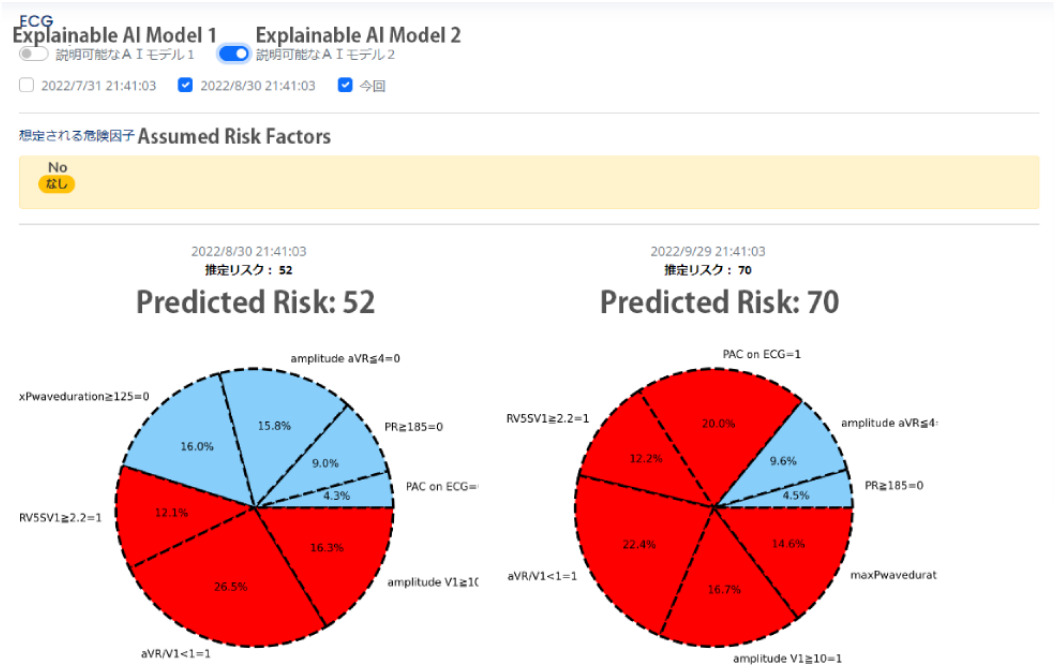


Figure 5. Screenshot of the explainable items and how they were displayed in the system—explainable model 2. aVR: Augmented Vector Right; ECG: electrocardiography. PAC: Premature Atrial Contraction; PR: PR Interval (the onset of the P wave to the start of the QRS complex).



Study Process

Each cardiologist participated in a 1-on-2 interview with an HCI researcher and a cardiologist. Before the interview began, the HCI researcher set the scene for the cardiologist using the following statement: "You have 11 patients that you are currently caring for. You can see a list of them in the AF'fective system, which indicates their risks levels for recurrence." Participants were first briefed on the system and then asked to test it by imagining the following four scenarios: (1) viewing a patient's status, (2) checking a patient's ECG graph, (3) viewing explainable model 1, and (4) viewing explainable model 2. We then asked the cardiologists to complete the study measurements (UEQ, TAM, and MARS, as described in a previous section)

and began the semistructured interviews. Each interview lasted 45 minutes to an hour. All the interviews were conducted in Japanese and fully recorded, including screen recordings of the participants testing the AF'fective PWA.

Analysis

For the survey data, we calculated the mean and SD for each item in all surveys except for the UEQ, which has its own analysis tool in Microsoft Excel (Microsoft Corp) format. We compared our mean scores for the TAM and MARS against the theoretical average scores to determine whether physicians experienced any usability issues related to the interface design, functionality, information quality, or accessibility.

For the interview data, we reviewed and analyzed the transcripts following the thematic analysis framework proposed by Clarke et al [50]. The analysis was carried out by 3 researchers, 2 from the HCI field and 1 with a cardiology background. As the 2 HCI researchers were more familiar with the analysis method selected and had a long history of conducting thematic analysis, the initial formation of the themes was first conducted by them. The HCI researchers first read through all the transcripts to familiarize themselves with the content and highlighted interesting quotes. The quotes were then summarized and independently interpreted as "topics of interest" by the HCI researchers. The researchers then grouped the topics into several potential "themes" and involved the cardiologist to approve and edit the themes. The cardiologist researcher then joined the evaluation to come up with the final findings that were deemed meaningful in both the HCI and medical fields.

Results

Overview

In the final stage, we recruited 7 male cardiologists (aged 31 to 55 years) with 8 to 30 years of cardiology experience. After the interviews, they completed the surveys, and we calculated the mean scores to identify potential usability or design issues. The results suggested no significant usability problems, and the prototype showed strong potential to motivate physicians' intention to use it due to its ease of use, reliable information, and explainable functionality. The detailed survey results are reported in the following sections.

UEQ Results: Positive Perspicuity

The results from the UEQ showed that perspicuity received a positive evaluation (mean 1.21, SD 0.96), whereas attractiveness (mean 0.56, SD 0.71), efficiency (mean 0.625, SD 0.27), dependability (mean 0.33, SD 0.27), stimulation (mean 0.5, SD 0.3), and novelty (mean 0.58, SD 0.27) received neutral evaluations. Perspicuity stood out as the most positively perceived aspect of the UX, suggesting that ease of use and clarity are strong points of the AF'fective system. The neutral evaluations for the other aspects (attractiveness, efficiency, dependability, stimulation, and novelty) indicated that physicians had average or mixed perceptions of these factors.

TAM Results: Above-Average Technical Acceptance

In addition, the TAM measures showed that perceived usefulness (mean 4.22, SD 0.31), perceived ease of use (mean 5.13, SD 0.77), and behavioral intention to use (mean 4.77, SD 0.68) scores were all above the average (mean 3.5). The behavioral intention to use score indicated a strong intention to use the AF'fective system. These results suggest that the identified factors motivated users to engage with the system.

MARS Results: Perceived High Functionality and Information Reliability

Finally, the results of the MARS showed above-average scores for functionality (mean 4.25, SD 0.42), esthetics (mean 3.61, SD 0.68), information reliability (mean 3.38, SD 0.54), and perceived impact on practice (mean 3.33, SD 0.76). These results indicated that users perceived high functionality and reliable information from the AF'fective system, whereas they rated its esthetics and perceived impact on practice as somewhat mediocre. We were particularly interested in the physicians' high ratings for information reliability as these suggested that they felt confident in understanding and agreeing with the predictive model's decisions. Notably, this rating was achieved without physicians knowing the predictive model's performance. This serves as a strong indicator that insights derived from explainable models play a crucial role in physicians' adaptation to our technology. This was deemed acceptable for a system in its prototype phase.

Discussion

Principal Findings

Overview

Aligned with our positive survey results, cardiologists expressed strong acceptance of the explainable features and willingness to use the system during the in-depth interviews. They also praised the system's enhanced efficiency, comprehensibility, comparability, consistency, and trustworthiness:

> *I think it is useful to be able to express in a general, universal, numerical way what we used to say in the past, like, "This person is likely to have a recurrence," by using a risk score. I think it is useful.* [P04; male]

> *In clinical practice, we honestly don't have that much time to look at ECGs and scores in detail, so I think AI is a way to make it easier for us to get the numbers out.* [P07; male]

> *I think it is easy to pick up if you can tell from a quick glance at the list that a person in red is at higher risk of recurrence, or that a person in green has a low probability of recurrence.* [P06; male]

While most cardiologists in our study viewed the real data–backed explainable AI system positively, we identified potential concerns and (possibly) false expectations for its real-life implementation. During the interviews, cardiologists were asked to envision using the system in their daily routines, leading to a realistic discussion after 2 rounds of iterative design. This probing step was crucial for understanding the feasibility and challenges of implementing the system in clinical settings. Our findings are discussed in the following sections through 4 key themes.

Explaining the Symptoms Does Not Equal to Identifying the Causes

Perhaps due to our intentional mimicry of cardiologists' methods for identifying irregularities in ECG graphs, participants showed high acceptance of the predicted results after viewing our 7-item AF clinical explainable features. In fact, P04 gave a very positive endorsement of why the explainable items were necessary:

> *I am of course curious as to why the risk was higher, so the AI [predictive] model would be rather meaningless without a graph [with explainable AI items] that can explain it, or perhaps it would make me wonder why the model [makes such a decision].* [P04; male]

According to P04, the predictive model's results may fail to convince them without an explanation of why the model makes a particular decision. P04's reaction suggests that physicians find explanations based on the new clinical explainable features both agreeable and trustworthy. However, a concern emerged regarding the limitations of these features. A key request was, "Then what should I do?"—cardiologists wanted to understand the features influencing the risk score and actionable strategies to address them. This aligns with previous insights showing a preference for clear indicators and actionable steps. In one case, P05 even suggested using the system to nudge patients toward behavior change:

> *We are at a stage where only these numbers [of explainable items] are available. 40 or something like this is the only thing that comes out, so what should I do [with each of the items]?* [P04; male]
>
> *I feel that the numbers alone don't mean much, or don't seem to lead to a change in [patient's] behavior.* [P05; male]

It should be noted that physicians treated each explainable feature as a "treatable item," which was not possible with an ECG graph–based explainable model. This improvement suggests that converting graphical highlights into standardized feature explanations can foster the expectation that addressing each item will reduce the risk score. In our current phase, it is still too early to determine how to further link our features with treatment approaches; however, this does indicate an interesting and operationalizable future implementation for explainable models.

Hence, in our current system, while these features influence the predictive outcome, clinical causal insights should be interpreted with caution. Our 7-item explainable features are based on standard ECG wave interpretation, identifying AF symptoms but not directly indicating underlying causes such as excessive drinking or sleep apnea. As noted in the document by Shapley [51], estimating causal effects from explainable items can be misleading. Cardiologists should be cautioned about this limitation and guided to set appropriate expectations for decision-making.

A key lesson from our interviews was the need for better curation of explainable features in the model, especially in clinical contexts. While various approaches provide transparency in AI decision-making [31,39,52,53], it is crucial that the chosen features are relevant, operationalizable, and actionable for medical professionals. This focus ensures that AI predictions are both understandable and practically applicable, enhancing postablation patient care.

Window for Interpretation: When 0% and 100% Recurrence Risk Does Not Exist

While a predictive model can theoretically provide risk scores from 0% to 100%, in real life, extreme scores such as 0% or 100% are unattainable. Participants noted that the possibility of such extreme scores diminishes the perceived realism and reliability of the model's predictions:

> *Theoretically speaking, it has to be 0, so even if it's low, it's still bad. You can't say 30 is a good thing. There is still a chance for a recurrence...So there are cases where it comes out as 0? (Cardiologist interviewer: That would be considered extreme score. Well, some might be zero, but I think the smallest is probably 3 or 4 [in the natural clinical contexts].)* [P03; male]
>
> *I'm not quite sure what [the risk score] means. There are a lot of questions about what 12 means and what we should think about it. It's not 0 [even for a normal person], So are you saying that AF can happen to anyone? (Interviewer: I don't think we can lower that number any further.)* [P05; male]

In a similar manner, participants also questioned whether a predictive score of 100% was at all meaningful in clinical practice:

> *Are you saying that even if all of [the items] were full, it would not be 100?(Interviewer: Even if [all of the items were] full, I think it will probably be around 80-92, which is the highest risk score [in the natural clinical contexts].)* [P01; male]

It should be noted that it is uncommon for a predictive model to yield a full 100% score in clinical contexts, particularly in postsurgery contexts. The contributing factors would generally contain complex interactions and dependencies.

Cardiologists emphasized that recurrence risk scores should be viewed as reference points, not absolute indicators. Establishing a threshold, such as 50, is essential for guiding actions, but borderline scores (eg, 49 or 51) still warrant patient warnings or follow-ups. Similar "borderline windows" were noted in a grief-related study [52] where psychologists suggested taking the score as a reference for concerning cases. Although the width of the borderline window can also be highly subjective for each individual cardiologist and possibly for patients, this approach is common in clinical practice. Future studies should develop nuanced protocols for borderline cases to prevent misdiagnoses and patient stress, potentially standardizing risk level interpretation, as suggested by P05:

> *I think it would be better to indicate the risk of developing the disease on a 5-point scale, such as low, medium, and high, because I don't think there is such a difference between 40 and 41.* [P05; male]

In treatment decision-making, another notable implication from our findings was the patient's presumptive belief that the risk score should be 0%, particularly after undergoing ablation. It was apparently conflicting with what the predictive model would reveal:

> *However, if you think about it, after AF ablation, the number of people who think that AF recurrence is quite small. For example, more than 70% to 80% of the patients think it is zero.* [P03; male]

Even cardiologists familiar with the postablation risk might expect that addressing the explainable items (see the Methods section for more discussion on these items) could eliminate the recurrence risk:

> *Considering the importance of risk scoring...I am very interested in the explanatory factors or items involved, and which of them are [positively correlated to the risk score]. If the purpose of this system is to follow up the patient, it would be even better if we could get such risk factors and know what kind of intervention is being done here.* [P01; male]

It struck the research team that popular explainable AI approaches could lead to misinterpretations and unrealistic expectations, such as believing that a 0% recurrence risk is easily attainable. Our study highlights the need to set accurate expectations about what predictive results and explainable features truly imply in a clinical context and how much they can be influenced through intervention.

Offer a Holistic Risk Overview by Integrating Other Risk Factors and Temporal Data

In our study, we adopted the flow of cardiologists' diagnostic process by reviewing ECG graphs and assessing the presence of known risk factors one by one. Therefore, other known risk factors, such as high blood pressure, smoking, and drinking, were included as additional "risk factors" alongside the predicted risk score derived from the patient's ECG data. While we believed that this approach resembled a cardiologist's normal decision-making procedure and would increase their acceptance of the model's predicted results, our interviewees expressed concerns about the potential for the predictive model to yield biased risk estimates:

> *There are many known risks that are commonly associated with AF, such as high blood pressure, obesity, lack of exercise, etc... Inputting it would further make this risk [score], or perhaps, the evaluation, more correct.* [P03; male]

> *So we are predicting the risk score purely based on ECG...If you are going to make progress, I think it would be good to include obesity information, bmi, waist circumference, and so on.* [P06; male]

Furthermore, some of our participants also indicated that they would prefer a clear display and association with the medications or ongoing treatments that the patients are receiving. For instance, P02 requested us to explore whether the explainable model could reflect the effects of the patient's medications, and P01 indicated that it would be better to see which interventions were done to lead to the risk score:

> *If there was information on arrhythmia medications, it would be more useful. I think the doctor might be able to think about it more holistically, such as, "This ECG variation [highlighted by the explainable model] is caused by the medicine the patient is taking."* [P02; male]

> *If the purpose of this system is to follow up the patient, it would be even better if we could get such risk scores and know what kind of intervention is being done here.* [P01; male]

As the cardiologists pointed out, although our model provided an overview of the other risk factors, it would only be meaningful if they could establish a connection between the variation in risk scores and the specific risk factors. Our findings indicated that, for explainable elements to be truly useful in real-life practice, they need to relate to more holistic and treatment-related risk factors or medications. Otherwise, the contribution of such a predictive, even explainable, model to preventing disease recurrence would be quite limited.

In the end, perhaps the most important takeaway from the interviews was the need to provide a holistic overview and explanation of intrapatient temporal data and risk scores. Echoing the insights from previous interviews, the design strategy of comparing interpersonal ECGs (strategy G) was heavily criticized by the physicians as individual risk factors can vary significantly. However, comparing a patient's current and historical risk factors could provide valuable insights into the evolution of these factors and potential trends in recurrence risk. For instance, participants P04 and P06 described how they would like to plan their interventions by tracking variations in risk scores and identifying the underlying causal factors over different time points:

> *For a patient with a high estimated risk this time, but lower estimated risk in the previous times, why did it become so high? What factors contributed to it? [I would expect] this AI model to estimate and explain.* [P04; male]

> *I think it would be better to look at it...from a longitudinal point of view...to see what kind of intervention I should use, what's going wrong, and what the reason is for the increase [of the risk score].* [P06; male]

We were intrigued by the design opportunities revealed in our findings. As cardiologists reminded us, a significant and crucial aspect of their profession involves deciding on appropriate interventions. Although explainable AI models have shown promise in diagnosing and identifying patients at risk, they have yet to offer actionable insights for intervention planning. We believe that explainable AI tools should integrate comprehensive patient data, including temporal variations in risk factors, medications, and treatments, to enhance the model's relevance and utility in real-world settings. In addition, the insights generated should be clearly linked to potential intervention strategies that could influence risk scores. Without such integration, the potential of explainable AI tools to contribute meaningfully to patient care, particularly in preventing disease recurrence, may remain limited.

A Sense of Control and Learning From the Model

The findings of our interviews revealed that cardiologists valued the sense of control and the learning opportunities provided by these models, aligned with the findings in the previous rounds that suggested a self-actualizing and constant learning side of cardiologists.

One of the key insights that emerged from our study is the potential of explainable AI models to help cardiologists prioritize patients at higher risk. As P04 addressed, there was a potential for these models to become an integral part of daily practice:

> *If it becomes a habit, I will probably look at people with high estimated risk and wonder why they were at higher risk.* [P04; male]

Moreover, P02 also offered a similar thought regarding triaging patients based on the risk scores:

> *So the follow-up should be more thorough, and those who are more likely to be affected [by the recurrence risk] should be followed up in a shorter period of time.* [P02; male]

By discriminating the patients that need urgent attention, physicians gain a sense of control over who to allocate more time and treatment resources to. Cardiologists indicated that such a potential new routine would enhance their work efficiency and allow them to spare cognitive capacity for more critical treatment planning. Interestingly, our interview findings suggest that cardiologists may appreciate useful "hints" or "interpretations" from explainable models to help them narrow down potential problems:

> *I understand that the ECG is used as the reference, but I think it would be easier to understand if there was an explanation of what is of interest in this ECG, even though those who understand may [already] understand it.* [P01; male]

The power dynamic between cardiologists and the AI model is worth exploring. To what extent should an AI model's interpretations be considered, and to what extent should cardiologists accept its suggestions? These boundaries are critical to explore to ensure that the final AI models can be effectively integrated into the patient care routines.

Further enhancement of control over the prevention procedure might also occur as cardiologists begin to learn more from AI-augmented interpretation and actionability. Interestingly, the explainability of AI models appears to increase cardiologists' desire to validate their treatment approaches:

> *So, at the time of the ablation, if I burn the tricuspid valve, you know [the score will go down], or if I burn the superior vena cava, you know [the score will go down]...you can learn more and more from time to time.* [P03; male]
>
> *I would take another ECG before prescribing too much medication, and see if [the risk score] goes up or down on the next ECG...For those in yellow...I will adjust my [treatment plan] according to the risk level of the next ECG.* [P06; male]

P06 went further and emphasized that, while AI models are valuable for providing quantitative assessments, their true worth lies in delivering actionable insights that can guide clinical decision-making. This sentiment reflects a broader interest among cardiologists in understanding how AI models arrive at their conclusions, particularly in relation to identifying intervention points:

> *I wonder if you are highlighting these risk factors to tell us these are the areas for intervention. I wonder if there is a reason for this risk factor, such as if we treat these factors, it could lower the risk.* [P06; male]

In general, our findings indicate that cardiologists, particularly self-actualizers, are not only open to integrating AI suggestions into their decision-making processes but also eager to enhance their knowledge through interaction with the AI model provided that these suggestions are accompanied by clear rationales and practical recommendations.

Implications of Creating a Software as a Medical Device

In the development of AI algorithms for clinical applications, it is crucial to not only meet the user interface requirements but also ensure compliance with regulatory standards. Software as a medical device must adhere to international standards such as International Organization for Standardization (ISO) 13485 (medical device quality management systems), ISO 14971 (medical device risk management), and ISO 62304 (software life cycle processes). In addition, compliance with Food and Drug Administration approval and the European Medical Device Regulation is essential from the perspectives of patient safety, effectiveness, and legal requirements.

While the AF'fective study provides valuable foundational knowledge on user interface, it is still considered in the early stages of technology readiness level (1-3). However, with advancements in AI technology, the insights from this research could contribute to the revision of existing standards, such as the International Electrotechnical Commission 62366-1:2015 (usability engineering for medical devices). Specifically, the consideration of explainability in the AI result display methods proposed in this research—such as visually clarifying the rationale behind suggestions and intuitively conveying risk and uncertainty in interface design—will play a crucial role in helping health care professionals accurately understand AI-generated proposals and integrate them into clinical decision-making. Moreover, our study highlighted that health care professionals do not need to fully comprehend the technical background or mechanisms of AI. Instead, it is crucial to create a state in which users can reasonably understand "why this result was reached," which is effective in building trust in AI. Such efforts extend beyond display design and have the potential to be incorporated into new standards that facilitate smooth communication between AI and health care professionals.

Conclusions

This paper reports a 3-stage study involving 23 physicians to co-design, assess, and pilot an explainable AI system for AF postablation monitoring called AF'fective. Our co-design approach effectively identified 4 feasible explainable strategies.

Furthermore, while we actively used explainable AI to identify the features contributing to the predictive results, our explainable strategies enabled the design of a more holistic patient monitoring system that incorporates contextual factors such as patient history and established principles to support physicians' decision-making rather than merely explaining the features indicated by ECG graphs. In the third stage, the system prompted cardiologists to envision its use in their routines, leading to highly contextualized feedback. This study highlighted the need to curate actionable explainable features and set correct expectations for interpreting predictive scores. In addition, cardiologists were interested in understanding the AI's reasoning and identifying strategies to address recurrence risk factors.

Interestingly, although our focus was on AF recurrence prevention, many insights appear generalizable to other disease care contexts, such as temporal risk score monitoring and setting realistic expectations for scoring, particularly as extreme scores (0% and 100%) are practically unattainable. Further investigation is needed to determine whether these insights can be universally applied to enhance the feasibility of implementing explainable AI in clinical settings.

A future study will build on these findings and further evaluate the potential challenges and advantages when implementing the AF'fective system in near-live or live clinical care routines. We believe that AI possesses the potential to revolutionize medical practices and we can only realize it through putting the technology in actual use case scenarios.

Acknowledgments

This work was supported by a research grant from the Japanese Circulation Society (2021).

Conflicts of Interest

None declared.

References

1. von Eschenbach WJ. Transparency and the black box problem: why we do not trust AI. Philos Technol 2021 Sep 01;34(4):1607-1622. [doi: 10.1007/S13347-021-00477-0]
2. Petch J, Di S, Nelson W. Opening the black box: the promise and limitations of explainable machine learning in cardiology. Can J Cardiol 2022 Feb;38(2):204-213 [FREE Full text] [doi: 10.1016/j.cjca.2021.09.004] [Medline: 34534619]
3. Bizopoulos P, Koutsouris D. Deep learning in cardiology. IEEE Rev Biomed Eng 2019;12:168-193. [doi: 10.1109/RBME.2018.2885714] [Medline: 30530339]
4. Nakamura T, Sasano T. Artificial intelligence and cardiology: current status and perspective. J Cardiol 2022 Mar;79(3):326-333 [FREE Full text] [doi: 10.1016/j.jjcc.2021.11.017] [Medline: 34895982]
5. Nattel S. New ideas about atrial fibrillation 50 years on. Nature 2002 Jan 10;415(6868):219-226. [doi: 10.1038/415219a] [Medline: 11805846]
6. Lippi G, Sanchis-Gomar F, Cervellin G. Global epidemiology of atrial fibrillation: an increasing epidemic and public health challenge. Int J Stroke 2021 Feb;16(2):217-221. [doi: 10.1177/1747493019897870] [Medline: 31955707]
7. Odutayo A, Wong CX, Hsiao AJ, Hopewell S, Altman DG, Emdin CA. Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis. BMJ 2016 Sep 06;354:i4482 [FREE Full text] [doi: 10.1136/bmj.i4482] [Medline: 27599725]
8. Lip GY, Tse HF. Management of atrial fibrillation. Lancet 2007 Aug 18;370(9587):604-618. [doi: 10.1016/S0140-6736(07)61300-2] [Medline: 17707756]
9. Dretzke J, Chuchu N, Agarwal R, Herd C, Chua W, Fabritz L, et al. Predicting recurrent atrial fibrillation after catheter ablation: a systematic review of prognostic models. Europace 2020 May 01;22(5):748-760 [FREE Full text] [doi: 10.1093/europace/euaa041] [Medline: 32227238]
10. Shashikumar SP, Shah AJ, Li Q, Clifford GD, Nemati S. A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. In: Proceedings of the 2017 IEEE EMBS International Conference on Biomedical & Health Informatics. 2017 Presented at: BHI '17; February 16-19, 2017; Orlando, FL p. 141-144 URL: https://ieeexplore.ieee.org/document/7897225 [doi: 10.1109/bhi.2017.7897225]
11. Andersen RS, Peimankar A, Puthusserypady S. A deep learning approach for real-time detection of atrial fibrillation. Expert Syst Appl 2019 Jan;115:465-473. [doi: 10.1016/j.eswa.2018.08.011]
12. Zhou X, Nakamura K, Sahara N, Takagi T, Toyoda Y, Enomoto Y, et al. Deep learning-based recurrence prediction of atrial fibrillation after catheter ablation. Circ J 2022 Jan 25;86(2):299-308 [FREE Full text] [doi: 10.1253/circj.CJ-21-0622] [Medline: 34629373]
13. Hung M, Lauren E, Hon E, Xu J, Ruiz-Negrón B, Rosales M, et al. Using machine learning to predict 30-day hospital readmissions in patients with atrial fibrillation undergoing catheter ablation. J Pers Med 2020 Aug 09;10(3):82 [FREE Full text] [doi: 10.3390/jpm10030082] [Medline: 32784873]

14. Przewlocka-Kosmala M, Marwick TH, Dabrowski A, Kosmala W. Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: a classification based on machine learning. J Am Soc Echocardiogr 2019 May;32(5):604-15.e6. [doi: 10.1016/j.echo.2018.12.002] [Medline: 30718020]
15. Mortazavi BJ, Downing NS, Bucholz EM, Dharmarajan K, Manhapra A, Li SX, et al. Analysis of machine learning techniques for heart failure readmissions. Circ Cardiovasc Qual Outcomes 2016 Nov;9(6):629-640. [doi: 10.1161/circoutcomes.116.003039]
16. Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In: Proceedings of the 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017 on Statistical Atlases and Computational Models of the Heart, ACDC and MMWHS Challenges. 2017 Presented at: STACOM '17; September 10-14, 2017; Montreal, QC p. 120-129 URL: https://link.springer.com/chapter/10.1007/978-3-319-75541-0_13 [doi: 10.1007/978-3-319-75541-0_13]
17. Nagarajan VD, Lee SL, Robertus JL, Nienaber CA, Trayanova NA, Ernst S. Artificial intelligence in the diagnosis and management of arrhythmias. Eur Heart J 2021 Oct 07;42(38):3904-3916 [FREE Full text] [doi: 10.1093/eurheartj/ehab544] [Medline: 34392353]
18. Ebrahimi Z, Loni M, Daneshtalab M, Gharehbaghi A. A review on deep learning methods for ECG arrhythmia classification. Expert Syst Appl 2020 Sep;7:100033. [doi: 10.1016/j.eswax.2020.100033]
19. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. JAMA Cardiol 2018 May 01;3(5):409-416 [FREE Full text] [doi: 10.1001/jamacardio.2018.0136] [Medline: 29562087]
20. Nazarian S, Lam K, Darzi A, Ashrafian H. Diagnostic accuracy of smartwatches for the detection of cardiac arrhythmia: systematic review and meta-analysis. J Med Internet Res 2021 Aug 27;23(8):e28974 [FREE Full text] [doi: 10.2196/28974] [Medline: 34448706]
21. Sehrawat O, Kashou AH, Noseworthy PA. Artificial intelligence and atrial fibrillation. J Cardiovasc Electrophysiol 2022 Aug 15;33(8):1932-1943 [FREE Full text] [doi: 10.1111/jce.15440] [Medline: 35258136]
22. Feeny AK, Chung MK, Madabhushi A, Attia ZI, Cikes M, Firouznia M, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. Circ Arrhythm Electrophysiol 2020 Aug;13(8):e007952 [FREE Full text] [doi: 10.1161/CIRCEP.119.007952] [Medline: 32628863]
23. Kim JY, Kim Y, Oh GH, Kim SH, Choi Y, Hwang Y, et al. A deep learning model to predict recurrence of atrial fibrillation after pulmonary vein isolation. Int J Arrhythm 2020 Nov 16;21(1):95. [doi: 10.1186/S42444-020-00027-3]
24. Budzianowski J, Kaczmarek-Majer K, Rzeźniczak J, Słomczyński M, Wichrowski F, Hiczkiewicz D, et al. Machine learning model for predicting late recurrence of atrial fibrillation after catheter ablation. Sci Rep 2023 Sep 14;13(1):15213 [FREE Full text] [doi: 10.1038/s41598-023-42542-y] [Medline: 37709859]
25. Li Z, Wheelock KM, Lathkar-Pradhan S, Oral H, Clauw DJ, Gunaratne P, et al. Predicting atrial fibrillation episodes with rapid ventricular rates associated with low levels of activity. BMC Med Inform Decis Mak 2021 Dec 28;21(1):364 [FREE Full text] [doi: 10.1186/s12911-021-01723-3] [Medline: 34963444]
26. Nishimura T, Senoo K, Makino M, Munakata J, Tomura N, Shimoo S, et al. Prediction model for the new onset of atrial fibrillation combining features of 24-hour Holter electrocardiogram with 12-lead electrocardiogram. Int J Cardiol Heart Vasc 2023 Aug;47:101245 [FREE Full text] [doi: 10.1016/j.ijcha.2023.101245] [Medline: 37521520]
27. Hallowell N, Badger S, Sauerbrei A, Nellåker C, Kerasidou A. "I don't think people are ready to trust these algorithms at face value": trust and the use of machine learning algorithms in the diagnosis of rare disease. BMC Med Ethics 2022 Nov 16;23(1):112 [FREE Full text] [doi: 10.1186/s12910-022-00842-4] [Medline: 36384545]
28. K S, V S, EA G, KP S. Explainable artificial intelligence for heart rate variability in ECG signal. Healthc Technol Lett 2020 Dec 09;7(6):146-154. [doi: 10.1049/htl.2020.0033]
29. Jo YY, Cho Y, Lee SY, Kwon J, Kim KH, Jeon KH, et al. Explainable artificial intelligence to detect atrial fibrillation using electrocardiogram. Int J Cardiol 2021 Apr 01;328:104-110. [doi: 10.1016/j.ijcard.2020.11.053] [Medline: 33271204]
30. Ma Y, Zhang D, Xu J, Pang H, Hu M, Li J, et al. Explainable machine learning model reveals its decision-making process in identifying patients with paroxysmal atrial fibrillation at high risk for recurrence after catheter ablation. BMC Cardiovasc Disord 2023 Feb 17;23(1):91 [FREE Full text] [doi: 10.1186/s12872-023-03087-0] [Medline: 36803424]
31. Anand A, Kadian T, Shetty MK, Gupta A. Explainable AI decision model for ECG data of cardiac disorders. Biomed Signal Process Control 2022 May;75:103584. [doi: 10.1016/j.bspc.2022.103584]
32. Mousavi S, Afghah F, Acharya UR. HAN-ECG: an interpretable atrial fibrillation detection model using hierarchical attention networks. Comput Biol Med 2020 Dec;127:104057 [FREE Full text] [doi: 10.1016/j.compbiomed.2020.104057] [Medline: 33126126]
33. Mousavi S, Afghah F, Razi A, Acharya UR. ECGNET: learning where to attend for detection of atrial fibrillation with deep visual attention. IEEE EMBS Int Conf Biomed Health Inform 2019 May;2019:854 [FREE Full text] [doi: 10.1109/BHI.2019.8834637] [Medline: 33083788]
34. Raza A, Tran KP, Koehl L, Li S. Designing ECG monitoring healthcare system with federated transfer learning and explainable AI. Knowl Based Syst 2022 Jan;236:107763. [doi: 10.1016/j.knosys.2021.107763]

35. Wesołowski S, Lemmon G, Hernandez EJ, Henrie A, Miller TA, Weyhrauch D, et al. An explainable artificial intelligence approach for predicting cardiovascular outcomes using electronic health records. PLOS Digit Health 2022 Jan 18;1(1):e0000004 [FREE Full text] [doi: 10.1371/journal.pdig.0000004] [Medline: 35373216]
36. Matthiesen S, Diederichsen SZ, Hansen MK, Villumsen C, Lassen MC, Jacobsen PK, et al. Clinician preimplementation perspectives of a decision-support tool for the prediction of cardiac arrhythmia based on machine learning: near-live feasibility and qualitative study. JMIR Hum Factors 2021 Nov 26;8(4):e26964 [FREE Full text] [doi: 10.2196/26964] [Medline: 34842528]
37. Gerlings J, Jensen MS, Shollo A. Explainable AI, but explainable to whom? An exploratory case study of xAI in healthcare. In: Lim CP, Chen YW, Vidya A, Mahorkar C, Jain LC, editors. Handbook of Artificial Intelligence in Healthcare. Volume 2. Cham, Switzerland: Springer; 2022:169-198.
38. Payrovnaziri SN, Chen Z, Rengifo-Moreno P, Miller T, Bian J, Chen J, et al. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. J Am Med Inform Assoc 2020 Jul 01;27(7):1173-1185 [FREE Full text] [doi: 10.1093/jamia/ocaa053] [Medline: 32417928]
39. She WJ, Senoo K, Iwakoshi H, Kuwahara N, Siriaraya P. AF'fective design: supporting atrial fibrillation post-treatment with explainable AI. In: Proceedings of the 27th International Conference on Intelligent User Interfaces. 2022 Presented at: IUI '22; March 22-25, 2022; Helsinki, Finland p. 22-25 URL: https://dl.acm.org/doi/10.1145/3490100.3516455 [doi: 10.1145/3490100.3516455]
40. Xie Y, Chen M, Kao D, Gao G, Chen XA. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-13 URL: https://dl.acm.org/doi/10.1145/3313831.3376807 [doi: 10.1145/3313831.3376807]
41. Alhadreti O, Mayhew P. Rethinking thinking aloud: a comparison of three think-aloud protocols. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-12 URL: https://dl.acm.org/doi/10.1145/3173574.3173618 [doi: 10.1145/3173574.3173618]
42. Laugwitz B, Held T, Schrepp M. Construction and evaluation of a user experience questionnaire. In: Proceedings of the 4th Symposium of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society. 2008 Presented at: USAB '08; November 20-21, 2008; Graz, Austria p. 63-76 URL: https://link.springer.com/chapter/10.1007/978-3-540-89350-9_6 [doi: 10.1007/978-3-540-89350-9_6]
43. Schrepp M, Hinderks A, Thomaschewski J. Construction of a benchmark for the user experience questionnaire (UEQ). Int J Interact Multimed Artif Intell 2017;4(4):40. [doi: 10.9781/ijimai.2017.445]
44. Davis FD, Venkatesh V. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. Int J Hum Comput Stud 1996 Jul;45(1):19-45. [doi: 10.1006/ijhc.1996.0040]
45. Davis F, Bagozzi R, Warshaw P. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989;35(8):1003 [FREE Full text]
46. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). PLoS One 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: 10.1371/journal.pone.0241480] [Medline: 33137123]
47. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
48. Yamamoto K, Ito M, Sakata M, Koizumi S, Hashisako M, Sato M, et al. Japanese version of the Mobile App Rating Scale (MARS): development and validation. JMIR Mhealth Uhealth 2022 Apr 14;10(4):e33725 [FREE Full text] [doi: 10.2196/33725] [Medline: 35197241]
49. Lundberg S, Lee SI. A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017 Presented at: NIPS '17; December 4-9, 2017; Long Beach, CA p. 4768-4777 URL: https://dl.acm.org/doi/10.5555/3295222.3295230
50. Clarke V, Braun V, Hayfield N. Thematic analysis. In: Smith JA, editor. Qualitative Psychology: A Practical Guide to Research Methods. Thousand Oaks, CA: Sage Publications; 2015:23-51.
51. Lundberg S. Be careful when interpreting predictive models in search of causal insights. Medium. URL: https://towardsdatascience.com/be-careful-when-interpreting-predictive-models-in-search-of-causal-insights-e68626e664b6 [accessed 2024-04-29]
52. She WJ, Ang CS, Neimeyer RA, Burke LA, Zhang Y, Jatowt A, et al. Investigation of a web-based explainable AI screening for prolonged grief disorder. IEEE Access 2022;10:41164-41185. [doi: 10.1109/access.2022.3163311]
53. Taniguchi H, Takata T, Takechi M, Furukawa A, Iwasawa J, Kawamura A, et al. Explainable artificial intelligence model for diagnosis of atrial fibrillation using Holter electrocardiogram waveforms. Int Heart J 2021;62(3):534-539 [FREE Full text] [doi: 10.1536/ihj.21-094] [Medline: 34053998]

Abbreviations

AF: atrial fibrillation
AI: artificial intelligence
aVR: augmented vector right
CNN: convolutional neural network
ECG: electrocardiography
Grad-CAM: Gradient-Weighted Class Activation Mapping
HCI: human-computer interaction
IRB: institutional review board
ISO: International Organization for Standardization
MARS: Mobile App Rating Scale
PWA: progressive web application
SHAP: Shapley Additive Explanations
TAM: technology acceptance model
UEQ: User Experience Questionnaire
UX: user experience
V1: first precordial lead

Edited by A Kushniruk; submitted 29.08.24; peer-reviewed by D Almonfrey, HJ Noordmans; comments to author 14.10.24; revised version received 27.11.24; accepted 05.01.25; published 13.02.25.

Please cite as:
She WJ, Siriaraya P, Iwakoshi H, Kuwahara N, Senoo K
An Explainable AI Application (AF'fective) to Support Monitoring of Patients With Atrial Fibrillation After Catheter Ablation: Qualitative Focus Group, Design Session, and Interview Study
JMIR Hum Factors 2025;12:e65923
URL: https://humanfactors.jmir.org/2025/1/e65923
doi:10.2196/65923
PMID:

Original Paper

Preadolescent Children Using Real-Time Heart Rate During Moderate to Vigorous Physical Activity: A Feasibility Study

Lincoln Lu[1], PhD; Danielle E Jake-Schoffman[2], PhD; Hannah A Lavoie[2], MS; Maedeh Agharazidermani[1], PhD; Kristy Elizabeth Boyer[1], PhD

[1]LearnDialogue Lab, Computer and Information Science and Engineering, University of Florida, Gainesville, FL, United States

[2]Exhale Lab, Health and Human Performance, University of Florida, Gainesville, FL, United States

Corresponding Author:
Lincoln Lu, PhD
LearnDialogue Lab
Computer and Information Science and Engineering
University of Florida
1889 Museum Road
Gainesville, FL, 32611
United States
Phone: 1 352 392 1133
Email: lincolnlu@ufl.edu

Abstract

Background: Given the global burden of insufficient physical activity (PA) in children, effective behavioral interventions are needed to increase PA levels. Novel technologies can help expand the reach and accessibility of these programs. Despite the potential to use heart rate (HR) to target moderate- to vigorous-intensity PA (MVPA), most HR research to date has focused on the accuracy of HR devices or used HR for PA surveillance rather than as an intervention tool. Furthermore, most commercial HR sensors are designed for adults, and their suitability for children is unknown. Further research about the feasibility and usability of commercial HR devices is required to understand how children may use HR during PA.

Objective: This study aimed to explore the use of a chest-worn HR sensor paired with a real-time HR display as an intervention tool among preadolescent children and the usability of a custom-designed app (Connexx) for viewing real-time HR.

Methods: We developed Connexx, an HR information display app with an HR analytics portal to view HR tracking. Children were recruited via flyers distributed at local public schools, word of mouth, and social media posts. Eligible participants were children aged 9 to 12 years who did not have any medical contraindications to MVPA. Participants took part in a single in-person study session where they monitored their own HR using a commercial HR sensor, learned about HR, and engaged in a series of PAs while using the Connexx app to view their real-time HR. We took field note observations about participant interactions with the HR devices. Participants engaged in a semistructured interview about their experience using Connexx and HR during PA and completed the System Usability Scale (SUS) about the Connexx app. Study sessions were audio and video recorded and transcribed verbatim.

Results: A total of 11 participants (n=6, 55% male; n=9, 82%, non-Hispanic White) with an average age of 10.4 (SD 1.0) years were recruited for the study. Data from observations, interviews, and SUS indicated that preadolescent children can use real-time HR information during MVPA. Observational and interview data indicated that the participants were able to understand their HR after a basic lesson and demonstrated the ability to make use of their HR information during PA. Interview and SUS responses demonstrated that the Connexx app was highly usable, despite some accessibility challenges (eg, small display font). Feedback about usability issues has been incorporated into a redesign of the Connexx app, including larger, color-coded fonts for HR information.

Conclusions: The results of this study indicate that preadolescent children understood their HR data and were able to use it in real time during PA. The findings suggest that future interventions targeting MVPA in this population should test strategies to use HR and HR monitoring as direct program targets.

(JMIR Hum Factors 2025;12:e58715) doi:10.2196/58715

KEYWORDS

smartphone app; physical activity; heart rate; wearable sensors; youth; commercial wearable device; Garmin; mobile phone

Introduction

Background

Physical activity (PA) levels are insufficient across most of the world's population [1], constituting a persistent problem in global health. PA engagement in the United States is reflective of global trends: only 25% of US children meet the minimum daily recommended amount of moderate- to vigorous- PA (MVPA) [2]. PA behaviors and insufficient PA developed during childhood often worsen during adolescence if not addressed through intervention [3]. Insufficient PA is also a risk factor for obesity, a condition related to various comorbidities [4] that can persist into adulthood [5,6]. Alarmingly, children who develop obesity at a young age are more likely to have obesity into adulthood [7]. Therefore, addressing obesogenic factors (eg, PA) during childhood has the potential to positively impact long-term overall health [8,9].

Effective behavioral interventions are essential to address insufficient PA, particularly before the significant decline that often occurs during adolescence, around the age of 13 years [10]. Novel technologies can help expand the reach and accessibility of PA promotion efforts. To date, most research leveraging commercial PA devices has focused on adult populations [11], although some studies focusing on children have examined step counts as a PA outcome [12]. Step count measures basic movement, but this metric insufficiently captures the intensity of movement and, therefore, may be inadequate for measuring or promoting MVPA.

In contrast to step count, heart rate (HR) holds great promise for measuring MVPA. Moreover, it is possible that novel technologies allowing children to view their HR in real time could foster PA awareness and engagement. However, this great potential is understudied; a recent systematic review found only two studies that examined the use of existing commercial sports-style HR sensors for use in preadolescent children [13]. The studies identified by the systematic review focused on whether children would wear an HR sensor on their chest paired with a wrist-worn HR display over a long time period [14] and the accuracy of commercial HR sensors on a preadolescent child's wrist against a medical-grade control [15]. Another recent study involved the use of chest-worn HR sensors paired with a wrist-worn HR display to monitor PA in adolescents and young adults. However, the study did not provide participants with education on HR or guidance on effectively using HR to achieve MVPA goals [16].

These studies demonstrate the accuracy of commercial HR sensors and displays for preadolescent children. The adoption of devices capable of displaying real-time HR information is growing in this age group, matching trends across the demographic spectrum. Recent statistics indicate wide adoption of smartphones: 97% of Americans own a smartphone [17], and ownership is equitable across racial and socioeconomic groups [18]. Statistics from a survey conducted in July 2019 reported that approximately 21% of the US population owned a smartwatch [19]. Smart device adoption is growing among preadolescent children in the United States; by their teenage years, 95% of children either have their own smartphone or have access to one through their family [20]. Despite the growing ubiquity of these devices, research has only focused on the technical ability of smart devices to surveille user HR [16]. Research should now be conducted to explore the use of real-time HR as an intervention tool to target specific MVPA goals.

Objectives

Pairing chest-worn HR sensors with existing smart devices in the form of smartphones and smartwatches is a novel opportunity that has not been sufficiently explored. However, providing real-time HR information without exploring preadolescent children's ability to understand and use HR in real time to effectively target MVPA would only replicate existing studies with new hardware [16]. Therefore, we provided participants with information about HR and how to target different HR zones during PA. This study primarily aimed to explore whether preadolescent children can use real-time HR information to target specific MVPA goals. A chest-worn HR sensor was paired to a smart device (eg, smartphone or smartwatch) that displayed real-time HR information. As the HR sensors used in this study were originally designed for adult use, a secondary objective of this study was to explore whether preadolescent children could feasibly use a chest-worn HR sensor paired with a smart device.

Another secondary objective of this study was to explore the usability of a custom app (Connexx) developed by the research team. As part of a broader program of research, the team is exploring the impact of collaboration on PA, and one facet of this research explores collaborative pairs in a family dyad consisting of a parent or guardian and child. Dyads collaborate by viewing each other's real-time HR information during PA. There are currently no commercial apps capable of transmitting and displaying real-time HR information between two users. We developed the Connexx app with this functionality. However, a key aim of this study was to explore the usability of Connexx for preadolescent children, who represent half of the intended pair in parent or guardian and child collaborative dyads. Thus, this study explored the usability of the Connexx app in solo mode to establish its feasibility in use with preadolescent children.

Methods

Study Setting

This feasibility study was conducted at the University of Florida, Gainesville, Florida. Data collection occurred between October 2023 and February 2024.

Ethical Considerations

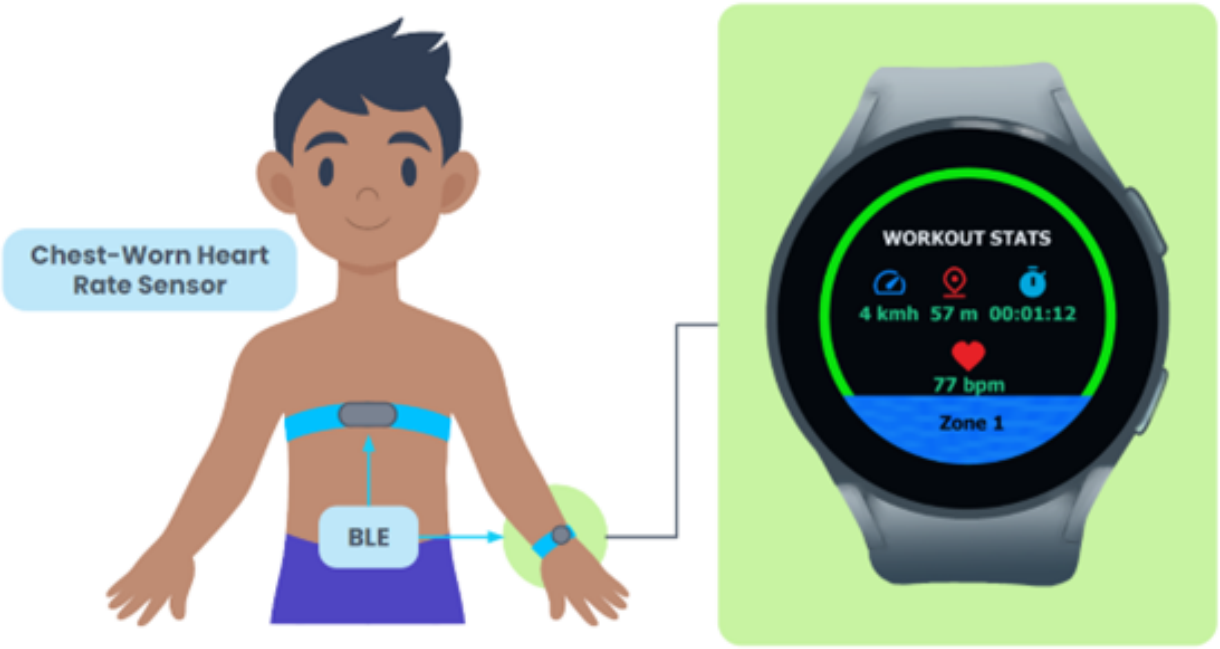
The study was approved by the University of Florida Institutional Review Board (IRB202301463) and complies with the ethical guidelines mandated therein. Informed consent was

obtained from a parent or guardian electronically, and written assent was obtained from the participants themselves at the start of each study session. Audio and video recordings from study sessions were stored on secured servers mandated as part of the University of Florida Institutional Review Board parameters. A US $25 Amazon e-gift card was provided to the participant's parent or guardian at the conclusion of the study session.

Study Equipment

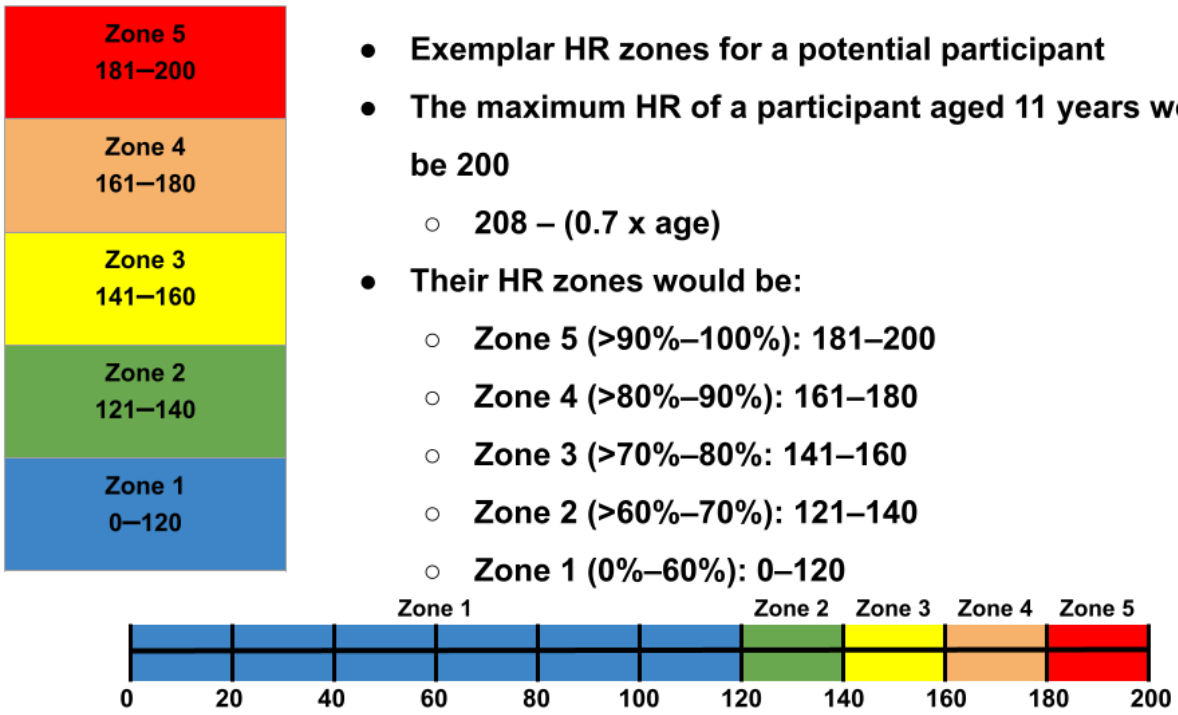
Several pieces of commercial technology were used to measure and display HR information (Figure 1). HR was measured using a sport-style chest-worn HR sensor (Garmin HRM-Pro Plus), and HR information was transmitted via Bluetooth Low Energy and displayed on either a smartphone or smartwatch. The smartphone used was a Google Pixel 4A, selected due to its low cost and the ability of Android phones to sideload apps, allowing us to install the Connexx app on the smartphone. The smartwatch used was a Samsung Galaxy Watch 4, selected for its Android-based operating system, which allows the Connexx app to be modified from the smartphone version to be compatible with the Galaxy Watch 4.

Figure 1. Smartwatch with solo workout screen connected to Bluetooth heart rate sensor. bpm: beats per minute; BLE: Bluetooth Low Energy.



This study used HR rather than accelerometry readings to measure MVPA. HR was chosen because it can be used to display effort in real time across various types of activities, which can contribute to a more accurate measure of the amount of time spent in MVPA [16]. We developed the Connexx app to display real-time HR information between two users. The Connexx app transmits real-time HR information between two devices via a traditional Bluetooth connection. However, Connexx is also capable of operating in solo mode, displaying the real-time HR of a single user, as we did in this study (Figure 1). Information on the user's current speed, distance covered during the workout, and workout duration was displayed along the top row of the Connexx solo workout screen. Speed and distance fields can be toggled between metric and imperial units by tapping on the display. The middle row of the workout screen displays the user's real-time HR. Connexx is able to display real-time HR information as either beats per minute (BPM) or as a percentage of maximum HR; users can toggle between these two display options by tapping on the HR display. A color-coded HR zone indicator is displayed below the HR field. This color coding matched the colors on HR exemplar cards used to explain HR information to participants (Figure 2). In this study, maximum HR was calculated using an established formula that is accurate for preadolescent children: 208 – (0.7 × age) [21]. The Connexx app automatically calculates HR zones using a straightforward 5-zone division based on the percentage of maximum HR, selected because there was no clear consensus or definitive definition of specific HR zones for preadolescent children [22-24] (zone 1: 0%-60%, zone 2: >60%-70%, zone 3: >70%-80%, zone 4: >80%-90%, and zone 5: >90%-100%). Furthermore, the Connexx app has an analytics portal; after the app records a workout, the information is sent to this portal, capable of displaying data from the workout on a graph.

Figure 2. Exemplar heart rate (HR) zone explanation card for a participant aged 11 years.



Participants, Eligibility, Recruitment, and Screening

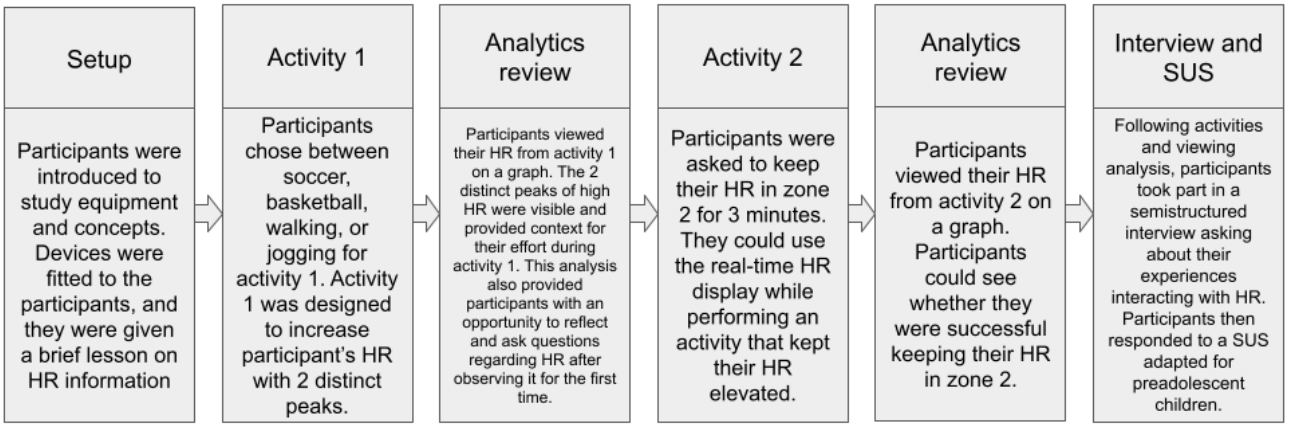
This study aimed to enroll a sample of 12 diverse preadolescent children to complete the full protocol. This sample size was chosen based on best practices in usability testing, showing that a sample of 10 to 12 participants is able to identify most major user issues [25]. Children were eligible to participate if they were aged 9 to 12 years and in good health with no physical conditions, injuries, or other conditions precluding them from engaging in MVPA for up to 30 minutes (assessed via a modified Physical Activity Readiness questionnaire for everyone [26]). We sought an equal number of male and female participants and equal distribution between Black, Hispanic, and non-Hispanic White participants (the three largest racial or ethnic groups in the local area).

Recruitment targeted parents of potentially eligible children using an electronic flyer distributed via local public elementary and middle schools, social media posts, and word of mouth. The flyer provided basic information regarding the study, eligibility criteria, compensation (US $25 Amazon e-gift card), and a link or QR code to a Qualtrics screening questionnaire. Parents or guardians completed the eligibility questionnaire, which included demographic information about their child. If the child was eligible, the parent or guardian was contacted via email to sign an electronic consent form through REDCap (Research Electronic Data Capture; Vanderbilt University). Once consent was obtained, a study team member (LL) scheduled the child for a study session.

Study Session

Study sessions were conducted at an indoor multiuse facility at the University of Florida. The space measures approximately 1021 square meters (27.7 meters by 36.9 meters) and can be used for a variety of uses, including basketball and volleyball; it has a maximum capacity of 1140 people. The study session began with obtaining assent for participation from the preadolescent child and followed the format outlined in Figure 3.

Figure 3. Study session activities flowchart. HR: heart rate; SUS: System Usability Scale.



Device Setup

Following assent, participants were fitted with Rode wireless lapel microphones. A researcher (LL) explained the study equipment and demonstrated the functionality of the chest-worn HR sensor. Parents aided participants in putting on the HR sensor. The session was recorded with a video camera. Next, participants chose between a smartphone and a smartwatch to view their real-time HR information. The Connexx app was available to all participants on either a smartphone or a smartwatch. However, all participants chose to use the smartwatch. The participants were guided through the process of entering their study information, such as their research identification number, age, and maximum HR, into the Connexx app. A workout was then initiated on the app, allowing participants to view their HR in real time, with more HR information provided to them.

HR Information

HR and HR zones were explained to the participants using a real-time HR display in Connexx (Figure 1) and an exemplar HR information card (Figure 2) based on their age (Table 1). Additional language used to explain HR and HR zones is provided in Multimedia Appendix 1. The goal of the brief explanation was to frame the upcoming activities in a way that participants could understand in relation to their HR zones. The researcher encouraged the participants to ask questions during the HR explanation.

Table 1. Example of the 5-zone heart rate (HR) categorization and description for participants aged 11 years.

HR zones	BPM[a]	Zone effort category	Example activities associated with the zone	Description of the zone
Zone 5	181 to 200	Maximum	Very high effort activities, which are often repeated (eg, sprinting in games such as tag multiple times in a row)	Extremely difficult to hold a conversation and most people must wait until after this effort to speak
Zone 4	161 to 180	Vigorous	Vigorous activities, such as playing a sport (eg, soccer or basketball), strenuous hiking, running, or climbing flights of stairs continually at a brisk pace	Difficult to hold conversation and often can only squeeze in short 1- or 2-word responses around breathing
Zone 3	141 to 160	Endurance	Shorter vigorous activities or long endurance activities (eg, climbing 2 to 3 flights of stairs or walking or hiking for an hour)	Slight difficulty to hold a normal conversation and speech is negotiated around breathing
Zone 2	121 to 140	Low endurance	Watching an exciting program and slow walk for short distances (eg, walking to the kitchen and walking around the block at a slow pace)	Can hold normal conversations without thinking about breathing
Zone 1	≤120	Rest	Sleeping, sitting and reading a book, or watching television	Can hold normal conversations without thinking about breathing

[a]BPM: beats per minute.

Activities

Next, we outlined the structure of the study to participants (Figure 3). The study consisted of 2 activity intervals; activity 1 lasted for 10 minutes and activity 2 lasted for 3 minutes, each followed by a brief rest period during which participants could view their HR information via an analytics portal on a laptop.

Activity 1

This study primarily aimed to explore whether preadolescent children can use real-time HR information to target specific MVPA goals. Activity 1 was designed to provide participants with an opportunity to observe their HR as it elevated through several different zones aligning with MVPA, providing practical, experiential use of HR information for participants. For activity 1, participants chose from playing basketball or soccer, or walking or jogging around the gym space. A researcher moderated participant effort during activity 1 by speeding up or slowing down the activity to take participants through a range of HR values. During this activity, the researcher asked participants to report their real-time HR at regular intervals. Querying for HR regularly was intended to help participants notice how their HR responded to various activity levels. Activity 1 was designed to increase participants' HR through multiple HR zones twice. Participants viewed their HR through the Connexx app while engaged in PA. After completing activity 1, the researcher guided participants through ending the workout on the Connexx app, and then, participants viewed a graphical representation of their HR in a Connexx analytics portal. We guided participants through the HR observation process, drawing attention to how HR can vary at rest, how HR responds to effort levels during PA, and how the body feels at various HRs.

Analytics Reflection

Once a workout ended on the Connexx app, workout data were uploaded to the analytics portal. The analytics portal displayed the participant's color-coded HR zones in the background of the graph, with HR in BPM on the vertical axis and duration on the horizontal axis (Figure 4). The researcher pointed to the connection between the HR graph and PA level: the more intense the PA level, the steeper the HR graph line became. Furthermore, the researcher pointed to how variable HR was: when at rest, HR did not stay at a consistent BPM; rather, it varied up and down slightly. In addition to the HR timeline graph, the analytics portal also presents an HR zone breakdown, which is a bar graph showing the total cumulative time the participants' HR spent in each zone (Figure 5). Participants were encouraged to ask questions throughout the process of viewing their HR graph.

Figure 4. Sample heart rate graph following activity 1. bpm: beats per minute.

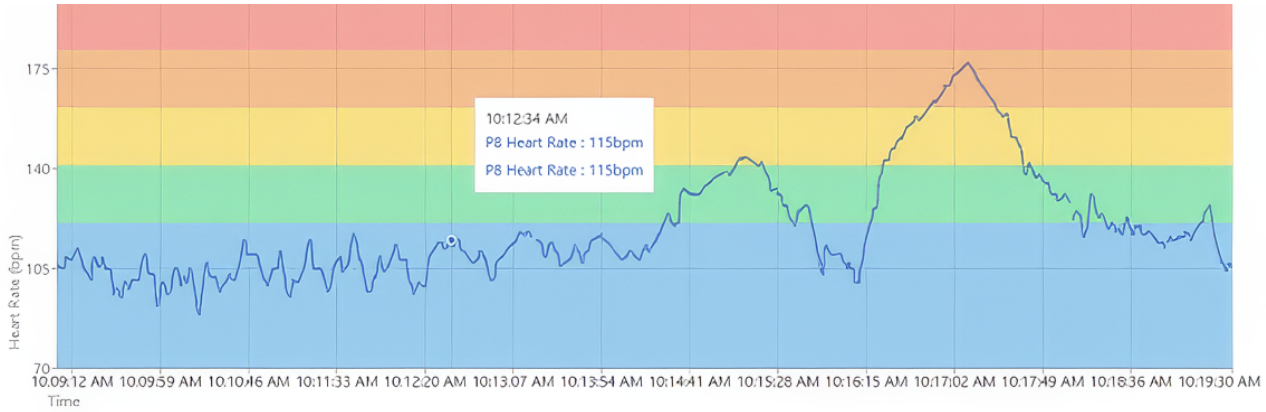
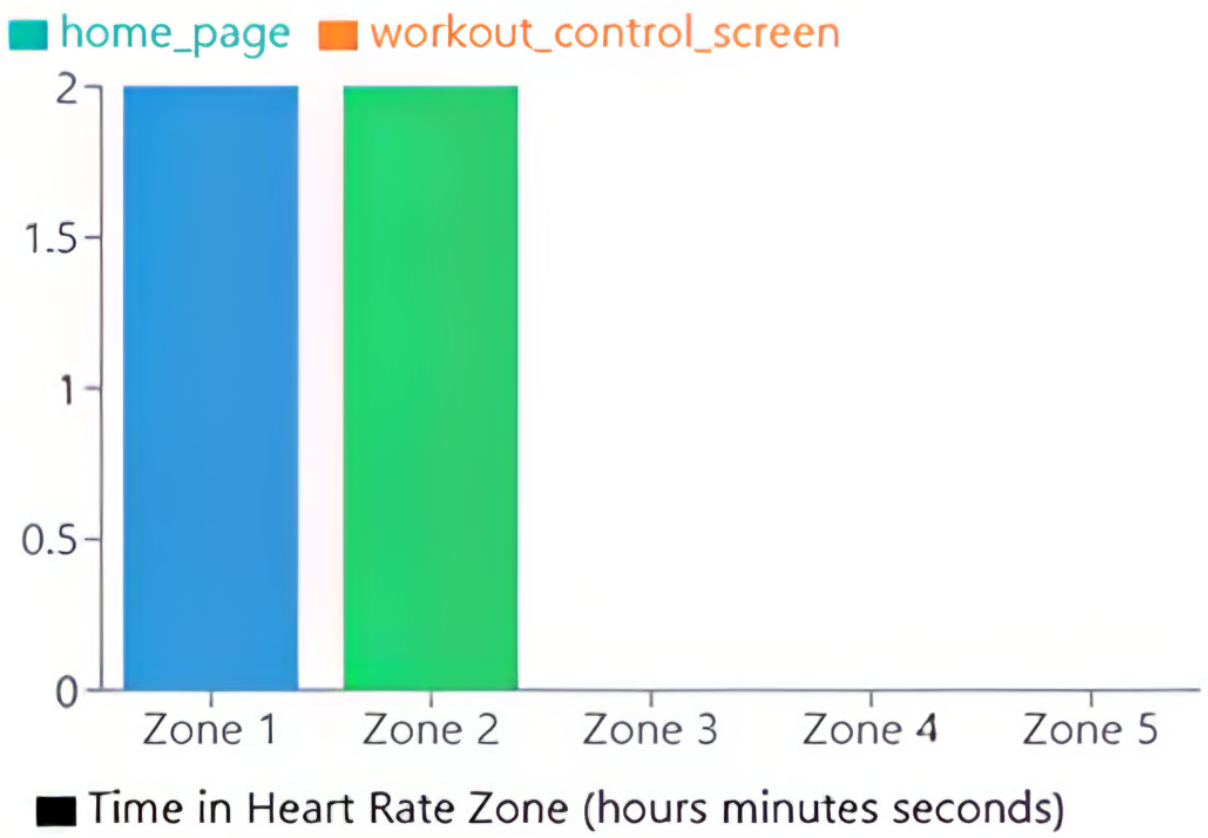


Figure 5. Sample heart rate zone cumulative time.



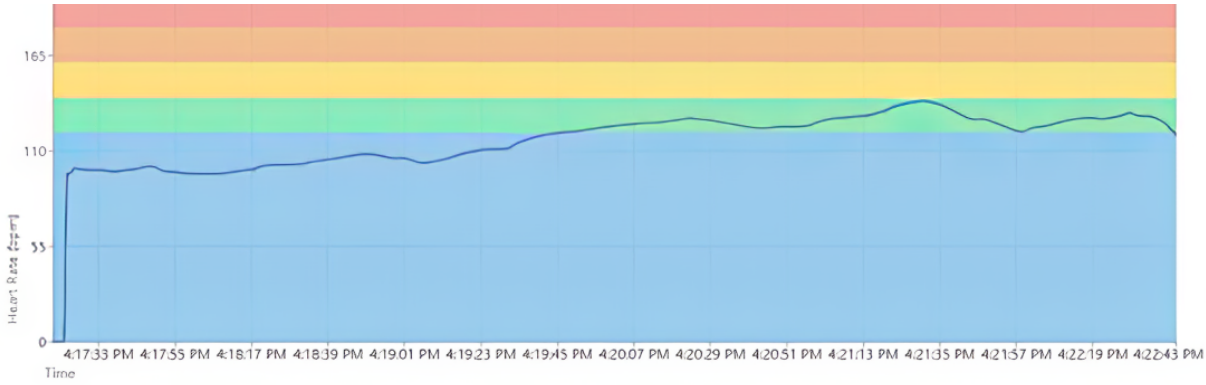
Activity 2

Activity 2 was designed to determine whether participants could make use of the tools provided to them so far (such as the information they have learned about HR, the HR sensor they were wearing, and the Connexx display, which provided a real-time display of their HR in BPM and HR zones) to maintain their HR in a targeted zone. Participants were asked to maintain their HR in zone 2 while engaged in the PA of their choice (basketball, soccer, walking, or jogging). Zone 2 was chosen because it was accessible to all participants, regardless of their fitness level, and they had to use the tools provided to maintain zone 2 HR, as they could easily exceed zone 2 HR by exercising too intensely. Participants could use the Connexx display to monitor their own HR and HR zone in real time to maintain zone 2 HR. The only prompts we provided were for timing (eg, when each minute elapsed). After 3 minutes, the participant was invited to save and end their workout on the Connexx app and view their HR graph for the second interval.

Analytics Review

The analytics portal displayed whether participants were successful in maintaining their HR in zone 2 (Figure 6). The HR zone breakdown also indicated the total time participants spent in each HR zone.

Figure 6. Sample heart rate graphical displays following activity 2. bpm: beats per minute.



Interview and System Usability Scale

After completing the 2 activities, participants engaged in a semistructured interview about their experiences. The interview explored participants' preference for smartwatch or smartphone, the comfort of commercial chest-worn HR sensors and smart devices, the participant's real-time actions to adjust their HR to meet a target goal, what HR zones mean to them, their preference for HR zones, whether they had previous experience with HR, and their enjoyment and potential future use of the Connexx app (refer to Multimedia Appendix 2 for the full interview script). During the interview, relevant objects and visual aids were presented to the participants as needed (eg, smartwatches and HR zone cards).

Finally, after the interview, participants filled out a modified version of the System Usability Scale (SUS) specifically modified and validated for use with preadolescent children [27]. This 10-item scale is used to calculate a usability score between 0 and 100. The commonly accepted threshold for a usable technology is a score of 68 [28].

Observational Data Collection Measures

We conducted observational data collection during each study session. In the event of recording capture device failure, a physical copy of the start and end times of each activity and a log of the participant's HR report and its corresponding value during the study session provide an overview of the participant's HR. Fortunately, this backup was not necessary during analysis.

The researcher who primarily interacted with the participants recorded field notes after each session. Notes focused on participant interaction with the instruments that might not be captured through the interview or SUS questionnaire, such as whether additional adjustments to the chest-worn HR sensor were required and how quickly participants adapted to the input methods for the Connexx app.

Analysis

Demographic data were summarized with descriptive statistics. Observational data were collated with the session transcripts to provide more depth and nuance to the thematic analysis. Additional observations that were not captured in the transcripts or SUS results are included as separate findings in the Observational Findings section.

Interviews were recorded and automatically transcribed via Microsoft OneDrive; we reviewed and revised autogenerated transcripts. Three researchers (LL, MA, and HAL) with qualitative research experience performed an inductive thematic analysis [29]. They followed the 6 steps of thematic analysis: data familiarization, initial code generation, searching for themes, theme review, defining and naming themes, and reporting [30,31]. While this study was designed with several objectives, we remained open to accepting any informative findings in the data. LL, MA, and HAL collaborated to familiarize themselves with the data and generate codes using a sample of the first 2 participant transcripts. The researchers worked together to search and identify themes, creating a codebook. The team then coded transcripts independently, meeting after completing the third and fourth participant transcripts to refine the codebook. Then, they coded the rest of the participant transcripts independently. Intercoder reliability was calculated using Krippendorff α [32] on a sample of 4 independently coded transcripts across 3 coders using the SPSS software (IBM Corp) extension developed by Hayes and Krippendorff [33]. Krippendorff α reliability estimate for the 3 coders was 0.907 (95% CI 0.8348-0.9621), corresponding to a high degree of agreement among them.

Results

Participant Demographics

A total of 11 preadolescent children with an average age of 10.5 (SD 1.0) years participated in this study, of whom 6 (55%) were male and 9 (82%) were non-Hispanic White. While the team sought to recruit a sample with equal numbers of male and female children and an even distribution of races and ethnicities, this diversity was not achieved in the final sample.

Observational Findings

Several observations demonstrated that preadolescent children can use real-time HR information to target specific MVPA goals. Participants shared a strong interest in their real-time HR information; some participants were quick to explore how their HR would respond to changes in PA. Participants looked at their HR information multiple times after initially starting a workout in the Connexx app and observed how controlling their breathing or fidgeting in their chairs may impact their HR. Their ability to use real-time HR information was further demonstrated by the actions of participants in activity 2. For activity 2,

participants were asked to maintain their HR in zone 2 for 3 minutes of PA. A total of 7 participants successfully used the Connexx display to hold their HR in zone 2. Of the remaining 4 participants, 1 (25%) initially ran at a high intensity that raised her HR above zone 2; however, throughout the rest of the 3-minute activity, she ran progressively slower to bring her HR back into zone 2. The 3 (75%) other participants took the opportunity to exert themselves as hard as possible in an attempt to achieve the highest HR reading instead of holding their HR in zone 2. These participants knew how to use the Connexx display to observe their HR in real time, as they looked at their smartwatch throughout activity 2. However, these 3 participants had a strong desire to see how high their HR could go. The parent of 1 of these 3 participants noted that she was highly competitive and always wanted to outcompete her older brother (who was also a participant in this study). One of the participants who successfully held his HR in zone 2 during the second activity demonstrated a similar desire to achieve a high HR value when he asked to be allowed to run around the gym space after completing the study so that he could see how fast his HR could read in the Connexx app.

Regarding hardware suitability for the children, all 11 participants chose to use a smartwatch to observe their HR in real time, and the watch straps successfully fit across a range of wrist sizes. However, there were some challenges with the chest-strap HR sensor. Two (18%) of the 11 participants needed the strap on the HR sensor to be shortened beyond its normal range to fit securely around their bodies. In both cases, the strap was shortened by <5 cm (by folding the excess fabric over and securing it with a safety pin) from the lower limit of the original strap range.

Furthermore, participants were able to quickly learn how to use the Connexx app. All participants were able to report their real-time HR and HR zone when prompted. Several participants (7/11, 64%) were also able to navigate the Connexx app themselves after the initial setup. They could swipe through the different screens in the app and start and end workouts themselves, and 2 (18%) out of 11 participants were able to navigate the text input interface with minimal guidance from the researcher. However, participants identified a substantial issue with the Connexx app in its current state. During the activities, participants noted that the font size of their HR display was too small. This was of particular concern during activity 1, as participants were moving vigorously at points to raise their HR through multiple HR zones. Participants had to adopt several different strategies to be able to see their HR values, including bringing the watch close to their face or slowing down their movement. The small font size of the Connexx app contributed to interruptions in the MVPA participants were engaged in.

Interview Findings

The interview provided rich data that generated themes related to using commercial devices to observe real-time HR to target specific MVPA goals. Three major themes emerged: one focused on real-time use of HR, one related to HR knowledge, and another on device and app use. Specific subthemes ranged from reports of real-time HR to preferences for the construction of HR zones. Themes, subthemes, definitions, and sample quotes are provided in Table 2. Major subthemes are described in the following sections in detail.

Table 2. Interview themes and sample quotes.

Theme and subtheme	Subtheme definition	Sample quotes
Real-time sensemaking of HR[a]		
Reporting HR and HR zones	Declarations of participant's real-time HR and HR zone	"Oh, I'm at Zone 4!... (which is) 160 (BPM)." [Participant 2]
Real-time HR use	How participants make use of real-time HR during activities	"I just looked at the watch a couple of times throughout the exercise and I saw Zone 2 and I'm like, OK, I'm, I was and I, you know, thought that I was doing good. So, I just kept it at that pace." [Participant 9]
HR and correlation to body sensations	How participant's body feels at different HR and HR zones	"I feel like I have to breathe more, but it mostly feels similar." [Participant 1]
HR-related knowledge		
Confusion regarding HR	Confusion about HR or HR zones	"I was just walking up and going over there to play soccer and it (HR) was (already) at 117." [Participant 4]
HR as an achievement	Desire for a high HR or HR zone displayed	"... with five, it's, I just like feel kind of accomplished." [Participant 8]
Previous knowledge of HR	Previous experience learning about HR	"In elementary school in PE class... count... how many beats we have in six seconds? And then he would tell us to, like, multiply it by 10 and that's beats per minute" [Participant 4]
HR zone preferences	Preference for HR zones, 3 larger or 5 smaller zones	"I personally like breaking stuff down to like smaller pieces and understanding it better." [Participant 9]
Device and app		
Device user experience	Comments on function and design of the device	"The smartwatch, you could just turn it over to look..." [Participant 7]
Device comfort or discomfort	How the devices felt for participants	"Wasn't too tight, too heavy... it feels comfortable." [Participant 3]
App usability	Like or dislike the app and willingness to use (or not use) the app in the future	"Probably just for like for fun, just to see (my HR)." [Participant 7]

[a]HR: heart rate.

Using Real-Time HR Information

During the interview, participants were asked about their strategies for adjusting their efforts to meet specific MVPA goals, including holding their HR in zone 2 for 3 minutes in activity 2. Participants who successfully held their HR in zone 2 discussed how they adjusted their effort based on the HR display (coded as "Real-time HR use" in Table 2):

> *I tried to keep the same speed when it was in Zone 2, because if I got faster, it [referring to his HR] will go higher, and if I went slower, it will go to Zone 1.* [Participant 3]

Often, this involved reducing their effort to not exceed the upper limit of zone 2 (coded as "HR and correlation to body sensations" in Table 2):

> *I tried using less power for [basketball] shots and like, try and walk or running across slower to conserve [energy].* [Participant 1]

The participants who were successful in keeping their HR steady in zone 2 used their smartwatch to adjust their effort much more often:

> *I just kept looking at the watch.* [Participant 2]

Some participants (2/11, 18%) noticed that their HR would change even when they did not think their actions should correspond to the change (coded as "Confusion regarding HR" in Table 2):

> *I was just walking up and going over there to play soccer and it [HR] was [already] at 117.* [Participant 4]

However, noting this unexpected result demonstrated that the participant was actively making connections between their PA and the displayed HR.

Not all participants were able to successfully keep their HR in zone 2 during activity 2. These participants acknowledged that they knew the goal was to maintain their HR in zone 2, but they wanted to achieve the highest HR possible. When asked what their HR was, participant 1 stated that "it just hit Zone 4, it is 161... I can go way faster than that!" (coded as "HR as achievement" in Table 2), and participant 8 expressed interest in his HR throughout the course of the study session. After finishing activity 1, he told the researcher that he "would like to see what it (his HR) looks like on a graph." Furthermore, after activity 2, where he was not successful in keeping his HR in zone 2, he told the researcher that he did "want to see how long I was in [zone] 5 for."

HR-Related Knowledge

In addition to findings on the specific research objectives, several themes emerged during the inductive thematic analysis process.

Previous Experience With HR

Four (36%) of the 11 participants had learned about HR in school before taking part in the study (coded as "Prior knowledge of HR" in Table 2); however, none of the participants had a complete understanding of HR. Learning about HR in school was summed up by a participant as follows:

> *[We would] do jumping jacks or jog in place for like 30 seconds, 20 seconds, and then... count our heart, how many beats we have in six seconds, and then he would tell us to, like multiply it by 10, and that's beats per minute.* [Participant 4]

None of the participants had used a chest-worn HR sensor before the study.

HR Zones

Participants were asked to describe what HR zones meant in their own words, and overall, they were able to provide a basic description of each zone. A participant explained the zones in a manner similar to how we had explained them before activity 1, as follows:

> *Zone 1, just calm, sitting down and not doing anything. Zone 2 I might be walking around, and I don't know. Zone 3 I would start to feel really hard to breathe and a little faster [breathing]. Zone 4, I might be breathing kind of heavier and having to slow down a little bit.* [Participant 4]

Other participants described the zones in relation to their own experiences during the study:

> *[Zone] 4 was like, I got to 4 when I was like shooting the ball and running.* [Participant 7]

Participants were asked whether they had a preference for the 5 HR zones used in this study or a more simplified 3 HR zones (coded as "HR zone preferences" in Table 2). Three (27%) of the 11 participants preferred the simplicity of having fewer zones:

> *It would be easier to look at if there was just easy, medium, hard. It would be easier to keep it in the easy-medium zone [referring to Zone 2].* [Participant 4]

The other participants (8/11, 73%) preferred 5 HR zones. Reasons for preferring 5 zones differed between participants; some preferred the greater granularity offered by more zones, and others preferred more HR zones as it provided a sense of achievement. A participant explained the meaning of having more HR zones as follows:

> *It would narrow it down and be easier just to keep it at that zone.* [Participant 6]

Another participant explained his preference for more HR zones as follows:

> *Oh, it's just like an achievement. It kind of feels like an achievement.* [Participant 7]
>
> *I just want to know if I'm in Zone 5, because, like, it's exciting.* [Participant 7]

Feasibility of Chest-Worn HR Sensors Paired to Smart Devices

Device Preference

To explore device preferences, participants were asked why they chose to use a smartwatch over a smartphone to observe their HR during PA (which all participants did). While only 2 (18%) of the 11 participants had consistently used a smartwatch before the study (one had their own, and another used their father's), all participants commented on the greater convenience they could foresee with a smartwatch, particularly while engaging in PA (coded as "Device user experience" in Table 2):

> *The smartwatch, you could just turn it over to look. But for a smartphone, you have to grab it, then you have to [mimes the action of pulling a phone out of their pocket to look at it]. If you're like working out, and you want to know [your HR] you could just look.* [Participant 7]

Device Comfort

When asked about the comfort or discomfort of the devices, all participants stated that the smartwatch felt comfortable (coded as "Device comfort or discomfort" in Table 2):

> *I'm used to wearing watches, so it's pretty normal for me to put one [smartwatch] on.* [Participant 1]

While there was some initial discomfort with the chest-worn HR sensor, it soon became unnoticeable:

> *When I first put it on, it kind of feels weird. But I got used to it when I used it for quite a while.* [Participant 3]

Only one participant reported continued discomfort and a desire to take off the HR sensor before the conclusion of the interview:

> *It was hot and kind of squeezing, so you get uncomfortable feeling.* [Participant 7]

Furthermore, one participant noted that using a softer fabric for the strap would increase the comfort of the instrument:

> *If like the material is different, so you didn't feel it as much. Like if it's like a smoother material or something maybe not as thick.* [Participant 6]

Usability of the Connexx App

Overall, participants were enthusiastic about the Connexx app and the ability to observe their own HR data (coded as "App usability" in Table 2). Participant 7 represented most participants when he declared that using Connexx "was kind of fun!" However, the small font size on the Connexx display made it challenging to view real-time HR while engaged in PA. For instance, participant 1 stated that what he "found frustrating was when we were running that when I wanted to check my HR, I couldn't see it straight because of it shaking." Many

participants (9/11, 82%) expressed interest in using the Connexx app in the future. The final response of participant 11 to a series of questions regarding the Connexx app and the ability to view her own HR was that she could "use it for playing games" with her friends. Similarly, participant 7 noted that he would like to use the Connexx app again in the future, "probably just for like for fun."

SUS Findings

The Connexx app can be considered highly usable in its current iteration; participants rated the Connexx app with an average score of 77.1 (SD 13.7), well above the threshold score of 68 [28]. However, feedback from this study will contribute to its future refinement to further increase usability.

Discussion

Principal Findings

While significant research has demonstrated the accuracy of commercially available HR devices for children [15], there has been insufficient research on how real-time HR information can be used to target MVPA goals. In this study, preadolescent children were provided with a basic lesson about their HR and HR zones, and then, they engaged in 2 PA sessions. Participants demonstrated an understanding of the HR display during PA by effectively using it to regulate their efforts. Therefore, real-time HR information can be used by children as a tool to meet PA guidelines more effectively; children can monitor their own HR information to ensure that they are engaged in MVPA. In addition, this study provides insight into the feasibility of HR sensors and smart devices, smart device preferences, the usability of a custom HR app, preadolescent children's previous HR experiences, and HR zone preferences.

Preadolescent Children Using Real-Time HR Information to Target Specific MVPA Goals

The findings of this study are unique in research on HR device use. Previous research often focused exclusively on accuracy [34], overlooked how users interact with commercial fitness trackers [35], or did not include HR information [36] in the design. In fact, preadolescent children in one study complained that the only real-time information they could see were step count and distance [36]. Participants in a study that paired chest-worn HR sensors with a wrist-worn display regulated their effort based on running speed; these participants did not use their available real-time HR information to adjust their effort [16]. Participants in this study demonstrated their ability to use real-time HR and HR zone information during PA by successfully keeping their HR in zone 2. We observed participants adjusting their effort while exercising (during activity 2) to keep their HR in zone 2. Participants confirmed in the interview that they adjusted their movement speed based on their real-time HR display, demonstrating an ability to use their real-time HR information to target specific MVPA goals during the study session.

Feasibility of Chest-Worn HR Sensors Paired to Smart Devices

Results of this study suggest that chest-worn HR sensors are feasible for use by preadolescent children. While some participants (2/11, 18%) mentioned slight discomfort when first putting on the chest-worn HR sensor, this did not seem to be an ongoing issue and was not brought up during activity 1 or 2. This study provides a more comprehensive investigation into the comfort and usability of these devices compared to previous research. One study that examined the use of a chest-strap HR sensor in conjunction with a wrist-worn device for preadolescent children focused on the accuracy of measurement rather than how preadolescent children felt about using these devices [14]. Another study focused on the accuracy of commercial devices for HR measurements, in this case, wrist-worn devices, without consideration for the comfort of the device [15]. Participants in this study were undergoing surgery, which may not be an appropriate situation to consider the comfort of an HR monitoring device. Furthermore, previous research that focused on the comfort of devices [16] examined an older population, whose more developed bodies may be better suited for devices typically designed for adults. However, the findings of this study indicate that chest-worn HR sensors and smartwatches were generally comfortable for preadolescent children.

Preference for Smartwatch

An interesting finding of this study on the feasibility of smart devices is the appeal of a smartwatch over a smartphone. Researchers did not anticipate all participants to choose a smartwatch over a smartphone for observing their HR information during the study. This is an issue that was not discussed in previous literature [37,38]. Upon reflection, this preference aligns with current commercial trends. Most preadolescent children (9/11, 82%) who participated in this study did not have previous experience with smartwatches or wristwatches in general, which may have contributed to a novelty effect. However, during the interview, they identified the convenience of using a wrist-worn device to view their HR in real time. Several participants (6/11, 54%) commented on perceived ease of use, where they would only have to move their wrists up to be able to view their HR rather than pulling a phone out of a pocket. Other participants (5/11, 45%) favored the security of a worn device, noting that they do not have to worry about dropping a watch compared to a phone. Furthermore, some participants (2/11, 18%) noted that watches may be more comfortable during PA, as a phone in their pocket would move undesirably.

Usability

Research on the design of mobile apps and the use of commercial wearable sensors has focused on the design process; technical limitations, such as loading speed; and efforts to ensure hardware capabilities and accuracy [39,40]. We developed the Connexx app following best practices based on industry standards and previous research [40,41]. Participants rated the Connexx app with a SUS score of 77, which is above the threshold score of 68 [28], indicating that the app has above-average usability. In addition, the Connexx app received a higher score than other digital health apps, more in line with

other PA apps [42]. The usability of the Connexx app was demonstrated by participants throughout the study.

For most participants (9/11, 82%), this was their first time using a smartwatch, and for all participants, this was their first time observing their real-time HR. We helped participants with the initial setup, such as inputting their age and other information in the Connexx app. However, many participants (7/11, 64%) quickly figured out how to operate the Connexx app themselves, and they could input text, navigate the menus, and begin and end workouts on their own without our intervention. This ability not only demonstrated the Connexx app's usability but also this population's familiarity with modern technology and its use, which was a concern for developing health-related apps for preadolescent children [43].

Despite the high SUS score, there were issues with the Connexx app. The Connexx app was initially developed for smartphones and then adapted to smartwatches, creating unanticipated usability challenges. The main PA screen for Connexx featured live tracking of several data fields, including HR, activity time, and speed. While this was easy to read on a smartphone display, on a smaller smartwatch display, the presence of multiple data fields resulted in each field being too small to be easily legible during PA. Thus, participants used multiple strategies to stabilize the smartwatch to be able to read what their HR was. The inclusion of a large color-coded portion at the bottom of the smartwatch display was helpful in informing participants of their HR zone at a glance. Despite difficulty reading portions of the app display, participants found it easy and fun to use. Furthermore, some participants (2/11, 18%) indicated that they not only wanted to use the app again in the future but also would like to share it with their friends.

Other HR Findings

Additional findings from this study indicated that an introductory lesson on HR helped provide participants with a basic understanding of their HR and demonstrated the feasibility of using HR during MVPA. Some participants (2/11, 18%) viewed high HR as an achievement, and the preferences for HR zones are linked to HR achievements for some participants.

Many participants (7/11, 64%) did not have previous experience with HR. An introductory lesson on HR and HR zones was provided for participants at the beginning of the study session, using a simple description of HR zones, including color-coded visual aids (Figure 2) and multiple descriptors that demonstrated how the participants' bodies might respond to various HR zones (Table 1). Participants subsequently demonstrated an understanding of their HR information while engaged in PA, responding to our queries about their HR with reports of both their HR and the corresponding HR zones. Furthermore, participants recalled the introductory lesson on HR well enough after engaging in PA to describe HR zones during the interview; some repeated the descriptors from the lesson, while others applied the descriptors to their own experiences. These findings suggest that participants were able to quickly grasp the HR concepts taught during the session and that using HR to target MVPA with preadolescent children is feasible.

This understanding of HR is even demonstrated by participants who were unable to keep their HR in zone 2 during activity 2. Participants who did not keep their HR in zone 2 demonstrated their understanding of HR by fully exerting themselves in an attempt to achieve a high HR reading. These participants were more interested in hitting the highest HR and HR zone possible; they viewed higher HR and HR zone as an accomplishment. While participants used the technology in a way not originally intended by the study's design, it still showed their understanding of HR and its application during PA.

Finally, many participants (8/11, 73%) preferred having a more granular HR zone division, favoring the 5-zone model over the 3-zone option. Some participants (2/11, 18%) equated zone 5 with a higher sense of accomplishment, mirroring the drive to achieve the highest HR possible. The drive for higher values is reflected in commercially available products for PA tracking. Platforms such as Strava [44] reward users by keeping track of their personal records. Furthermore, when a user's best performance is faster than any other user for a particular segment, the fastest user is rewarded with a crown icon and the title "King of the Mountain" or "Queen of the Mountain." The desire for high achievement of these participants is likely a reason why gamification and leaderboards are more effective in PA promotion than health benefits [45].

Limitations

Despite these findings, this study is not without its limitations. First, the participants had never encountered an app such as Connexx before; therefore, the novelty factor of a new measurement offered by the Connexx app may have offset potential shortcomings arising from font size or other use limitations [46]. As participants continue to use the Connexx app, the SUS score may decrease, as the novelty of displaying real-time HR information wears off. However, as participants indicated having little previous experience or education with HR, this highlights an educational opportunity missing from school curricula, as the participants were interested in learning about their HR.

In addition, the small sample size of this feasibility study limits the generalizability of the findings. While there were almost equal numbers of male and female participants, there was little racial or ethnic diversity (9/11, 82% non-Hispanic White). However, the sample size was still sufficient to demonstrate the usability of these HR monitoring technologies with preadolescent children [25]. Future research is needed to explore the technologies with a larger, more diverse sample.

Because this was a feasibility study to explore whether adolescent children could understand and use their HR information in real time as opposed to measuring HR change in response to PA, we did not control for other confounding variables (eg, the child's fitness level, diet, and resting HR). Rather, as a first step, this study simply examined whether HR use was possible. Moreover, due to the study design, this study did not take advantage of the full scope of capabilities developed as part of the Connexx app, including displaying a paired partner's HR information in real time. Future studies will further explore these capabilities of the Connexx app.

Conclusions

This study aimed to investigate the feasibility and usability of multiple technologies to measure and display HR information for preadolescent children. It examined whether preadolescent children could effectively use these devices to view their HR and make practical use of the information. The results indicate that preadolescent children understood and used their HR to regulate their MVPA and enjoyed using Connexx for real-time HR observation. Participant responses demonstrate the ability to use and the desire for access to real-time HR information during PA. Future research should build on these findings regarding the use of real-time HR as an intervention tool to target specific MVPA goals. This could involve testing strategies that use HR and real-time monitoring as direct targets in program designs. Preadolescent children's interest in using real-time HR should be leveraged by researchers, health professionals, and educators to promote MVPA in an engaging way for this population.

Acknowledgments

The authors would like to thank Sunny Dhama for assisting with software and investigation, Lidong Lin for assisting with the investigation, and Wesly Menard and Mollie Brewer for assisting with the conceptualization and methodology of the study. The funding was provided by the University of Florida College of Health and Human Performance and the University of Florida College of Engineering.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Explanation of heart rate zones.
[DOCX File , 14 KB - humanfactors_v12i1e58715_app1.docx]

Multimedia Appendix 2
Semistructured interview script.
[DOCX File , 10 KB - humanfactors_v12i1e58715_app2.docx]

References

1. Katzmarzyk PT. Expanding our understanding of the global impact of physical inactivity. Lancet Glob Health 2023 Jan;11(1):e2-e3. [doi: 10.1016/s2214-109x(22)00482-x]
2. Katzmarzyk PT, Denstel KD, Beals K, Bolling C, Wright C, Crouter SE, et al. Results from the United States of America's 2016 report card on physical activity for children and youth. J Phys Act Health 2016 Nov;13(11 Suppl 2):S307-S313 [FREE Full text] [doi: 10.1123/jpah.2016-0321] [Medline: 27848726]
3. Guthold R, Stevens GA, Riley LM, Bull FC. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. Lancet Child Adolesc Health 2020 Jan;4(1):23-35. [doi: 10.1016/s2352-4642(19)30323-2]
4. Finer N. Medical consequences of obesity. Medicine 2015 Feb;43(2):88-93. [doi: 10.1016/j.mpmed.2014.11.003]
5. Griffiths LJ, Parsons TJ, Hill AJ. Self-esteem and quality of life in obese children and adolescents: a systematic review. Int J Pediatr Obes 2010 Aug;5(4):282-304. [doi: 10.3109/17477160903473697] [Medline: 20210677]
6. Sahoo K, Sahoo K, Choudhury AK, Sofi NY, Kumar R, Bhadoria AS. Childhood obesity: causes and consequences. J Family Med Prim Care 2015;4(2):187-192 [FREE Full text] [doi: 10.4103/2249-4863.154628] [Medline: 25949965]
7. Obesity. World Health Organization. URL: https://www.who.int/health-topics/obesity#tab=tab_1 [accessed 2025-01-27]
8. Wyszyńska J, Ring-Dimitriou S, Thivel D, Weghuber D, Hadjipanayis A, Grossman Z, et al. Physical activity in the prevention of childhood obesity: the position of the European Childhood Obesity Group and the European Academy of Pediatrics. Front Pediatr 2020 Nov 5;8:535705 [FREE Full text] [doi: 10.3389/fped.2020.535705] [Medline: 33224905]
9. Dimitri P, Joshi K, Jones N, Moving Medicine for Children Working Group. Moving more: physical activity and its positive effects on long term conditions in children and young people. Arch Dis Child 2020 Nov 20;105(11):1035-1040. [doi: 10.1136/archdischild-2019-318017] [Medline: 32198161]
10. Davison KK, Werder JL, Trost SG, Baker BL, Birch LL. Why are early maturing girls less active? Links between pubertal development, psychological well-being, and physical activity among girls at ages 11 and 13. Soc Sci Med 2007 Jun;64(12):2391-2404 [FREE Full text] [doi: 10.1016/j.socscimed.2007.02.033] [Medline: 17451855]
11. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]

12. St Fleur RG, St George SM, Leite R, Kobayashi M, Agosto Y, Jake-Schoffman DE. Use of Fitbit devices in physical activity intervention studies across the life course: narrative review. JMIR Mhealth Uhealth 2021 May 28;9(5):e23411 [FREE Full text] [doi: 10.2196/23411] [Medline: 34047705]
13. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
14. Brazendale K, Decker L, Hunt ET, Perry MW, Brazendale AB, Weaver RG, et al. Validity and wearability of consumer-based fitness trackers in free-living children. Int J Exerc Sci 2019;12(5):471-482 [FREE Full text] [doi: 10.70252/BBTU4889] [Medline: 30899354]
15. Pelizzo G, Guddo A, Puglisi A, De Silvestri A, Comparato C, Valenza M, et al. Accuracy of a wrist-worn heart rate sensing device during elective pediatric surgical procedures. Children (Basel) 2018 Mar 08;5(3):38 [FREE Full text] [doi: 10.3390/children5030038] [Medline: 29518020]
16. Sun J, Liu Y. Using smart bracelets to assess heart rate among students during physical education lessons: feasibility, reliability, and validity study. JMIR Mhealth Uhealth 2020 Aug 05;8(8):e17699 [FREE Full text] [doi: 10.2196/17699] [Medline: 32663136]
17. Sidoti O, Dawson W, Gelles-Watnick R, Faverio M, Atske S, Radde K, et al. Mobile fact sheet. Pew Research Center. 2024 Nov 13. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2025-01-22]
18. Broffman L, Harrison S, Zhao M, Goldman A, Patnaik I, Zhou M. The relationship between broadband speeds, device type, demographic characteristics, and care-seeking via telehealth. Telemed J E Health 2023 Mar 01;29(3):425-431. [doi: 10.1089/tmj.2022.0058] [Medline: 35867048]
19. Vogels EA. About one-in-five Americans use a smart watch or fitness tracker. Pew Research Center. 2020 Jan 9. URL: https://www.pewresearch.org/short-reads/2020/01/09/about-one-in-five-americans-use-a-smart-watch-or-fitness-tracker/ [accessed 2025-01-27]
20. Richter A, Adkins V, Selkie E. Youth perspectives on the recommended age of mobile phone adoption: survey study. JMIR Pediatr Parent 2022 Oct 31;5(4):e40704 [FREE Full text] [doi: 10.2196/40704] [Medline: 36315219]
21. Tanaka H, Monahan KD, Seals DR. Age-predicted maximal heart rate revisited. J Am Coll Cardiol 2001 Jan;37(1):153-156 [FREE Full text] [doi: 10.1016/s0735-1097(00)01054-8] [Medline: 11153730]
22. Goh DY, Mun MW, Lee WL, Teoh OH, Rajgor DD. A randomised clinical trial to evaluate the safety, fit, comfort of a novel N95 mask in children. Sci Rep 2019 Dec 12;9(1):18952 [FREE Full text] [doi: 10.1038/s41598-019-55451-w] [Medline: 31831801]
23. Lind RR, Beck MM, Wikman J, Malarski K, Krustrup P, Lundbye-Jensen J, et al. Acute high-intensity football games can improve children's inhibitory control and neurophysiological measures of attention. Scand J Med Sci Sports 2019 Oct 24;29(10):1546-1562 [FREE Full text] [doi: 10.1111/sms.13485] [Medline: 31125468]
24. Miklós M, Komáromy D, Futó J, Balázs J. Acute physical activity, executive function, and attention performance in children with attention-deficit hyperactivity disorder and typically developing children: an experimental study. Int J Environ Res Public Health 2020 Jun 07;17(11):4071 [FREE Full text] [doi: 10.3390/ijerph17114071] [Medline: 32517384]
25. Alroobaea R, Mayhew PJ. How many participants are really enough for usability studies? In: Proceedings of the Science and Information Conference. 2014 Presented at: SAI 2014; August 27-29, 2014; London, UK. [doi: 10.1109/sai.2014.6918171]
26. Warburton DE, Jamnik VK, Bredin SS, Gledhill N. The physical activity readiness questionnaire for everyone (PAR-Q+) and electronic physical activity readiness medical examination (ePARmed-X+). Health Fit J Can 2011;4(2):3-17. [doi: 10.14288/hfjc.v4i2.103]
27. Putnam C, Puthenmadom M, Cuerdo MA, Wang W, Paul N. Adoption of the system usability scale for user testing with children. In: Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI EA '20; April 25-30, 2020; Honolulu, HI. [doi: 10.1145/3334480.3382840]
28. Lewis JR, Sauro J. Item benchmarks for the system usability scale. J Usability Stud 2018;13(3):158-167.
29. Sandborg J, Henriksson P, Larsen E, Lindqvist AK, Rutberg S, Söderström E, et al. Participants' engagement and satisfaction with a smartphone app intended to support healthy weight gain, diet, and physical activity during pregnancy: qualitative study within the HealthyMoms trial. JMIR Mhealth Uhealth 2021 Mar 05;9(3):e26159 [FREE Full text] [doi: 10.2196/26159] [Medline: 33666554]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. D'Addario M, Baretta D, Zanatta F, Greco A, Steca P. Engagement features in physical activity smartphone apps: focus group study with sedentary people. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e20460 [FREE Full text] [doi: 10.2196/20460] [Medline: 33196450]
32. Krippendorff K. Content Analysis: An Introduction to Its Methodology. Thousand Oaks, CA: SAGE Publications; 2019.
33. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. Commun Methods Meas 2007 Apr;1(1):77-89. [doi: 10.1080/19312450709336664]

34. Martinko A, Karuc J, Jurić P, Podnar H, Sorić M. Accuracy and precision of consumer-grade wearable activity monitors for assessing time spent in sedentary behavior in children and adolescents: systematic review. JMIR Mhealth Uhealth 2022 Aug 09;10(8):e37547 [FREE Full text] [doi: 10.2196/37547] [Medline: 35943763]
35. Chevance G, Golaszewski NM, Tipton E, Hekler EB, Buman M, Welk GJ, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022 Apr 13;10(4):e35626 [FREE Full text] [doi: 10.2196/35626] [Medline: 35416777]
36. Mackintosh KA, Chappel SE, Salmon J, Timperio A, Ball K, Brown H, et al. Parental perspectives of a wearable activity tracker for children younger than 13 years: acceptability and usability study. JMIR Mhealth Uhealth 2019 Nov 04;7(11):e13858 [FREE Full text] [doi: 10.2196/13858] [Medline: 31682585]
37. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298 [FREE Full text] [doi: 10.2196/mhealth.8298] [Medline: 31038460]
38. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR Mhealth Uhealth 2020 Nov 02;8(10):e20465 [FREE Full text] [doi: 10.2196/20465] [Medline: 33038869]
39. Düking P, Fuss FK, Holmberg HC, Sperlich B. Recommendations for assessment of the reliability, sensitivity, and validity of data provided by wearable sensors designed for monitoring physical activity. JMIR Mhealth Uhealth 2018 Apr 30;6(4):e102 [FREE Full text] [doi: 10.2196/mhealth.9341] [Medline: 29712629]
40. Newton A, Bagnell A, Rosychuk R, Duguay J, Wozney L, Huguet A, et al. A mobile phone-based app for use during cognitive behavioral therapy for adolescents with anxiety (MindClimb): user-centered design and usability study. JMIR Mhealth Uhealth 2020 Dec 08;8(12):e18439 [FREE Full text] [doi: 10.2196/18439] [Medline: 33289671]
41. Wu H, Li G. Innovation and improvement of visual communication design of mobile app based on social network interaction interface design. Multimed Tools Appl 2019 Apr 13;79:1-16. [doi: 10.1007/s11042-019-7523-6]
42. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
43. Schwarz A, Winkens LH, de Vet E, Ossendrijver D, Bouwsema K, Simons M. Design features associated with engagement in mobile health physical activity interventions among youth: systematic review of qualitative and quantitative studies. JMIR Mhealth Uhealth 2023 Mar 06;11:e40898 [FREE Full text] [doi: 10.2196/40898] [Medline: 36877551]
44. Community-powered motivation. STRAVA. URL: https://www.strava.com/ [accessed 2024-01-22]
45. Millonig A, Wunsch M, Stibe A, Seer S, Dai C, Schechtner K, et al. Gamification and social dynamics behind corporate cycling campaigns. Transp Res Proc 2016;19:33-39. [doi: 10.1016/j.trpro.2016.12.065]
46. Rutten I, Bogaert LV, Geerts D. From initial encounter with mid-air haptic feedback to repeated use: the role of the novelty effect in user experience. IEEE Trans Haptics 2021 Jul 1;14(3):591-602. [doi: 10.1109/toh.2020.3043658]

Abbreviations

BPM: beats per minute
HR: heart rate
MVPA: moderate- to vigorous- physical activity
PA: physical activity
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by A Kushniruk; submitted 22.03.24; peer-reviewed by H Namba, M Fraser; comments to author 01.12.24; revised version received 14.12.24; accepted 23.12.24; published 06.03.25.

Please cite as:
Lu L, Jake-Schoffman DE, Lavoie HA, Agharazidermani M, Boyer KE
Preadolescent Children Using Real-Time Heart Rate During Moderate to Vigorous Physical Activity: A Feasibility Study
JMIR Hum Factors 2025;12:e58715
URL: https://humanfactors.jmir.org/2025/1/e58715
doi:10.2196/58715
PMID:40053729

Original Paper

User-Centered Design of an Electronic Dashboard for Monitoring Facility-Level Basic Emergency Obstetric Care Readiness in Amhara, Ethiopia: Mixed Methods Study

Kylie Dougherty[1], RN, PhD; Yihenew Tesfaye[2], PhD; Heran Biza[3], BS, MS; Mulusew Belew[4], MSC; Natalie Benda[5], PhD; Abebe Gebremariam Gobezayehu[3,4], MD; John Cranmer[2,3,4,6,7], DNP, MPH, MSN, ANP, CPH, EBP-C; Suzanne Bakken[5], PhD

[1]Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]College of Health Sciences, Bahir Dar University, Bahir Dar, Ethiopia

[3]Nell Hodgson Woodruff School of Nursing, Emory University, Atlanta, GA, United States

[4]Emory-Ethiopia Partnership, Bahir Dar, Ethiopia

[5]School of Nursing, Columbia University, New York, NY, United States

[6]Woodruff Health Sciences Center, Emory University, Atlanta, GA, United States

[7]Center for the Study of Human Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Kylie Dougherty, RN, PhD
Department of Medical Social Sciences
Feinberg School of Medicine
Northwestern University
633 N St Clair St
Chicago, IL, 60611
United States
Phone: 1 (312) 503 1725
Email: kylie.dougherty@northwestern.edu

Abstract

Background: Maternal mortality remains a persistent public health concern in sub-Saharan African countries such as Ethiopia. Health information technology solutions are a flexible and low-cost method for improving health outcomes with proven benefits in low- to middle-income countries' health systems.

Objective: This study aimed to develop and assess the usability of an electronic dashboard to monitor facility-level readiness to manage basic emergency obstetric care (BEmOC) in Amhara, Ethiopia.

Methods: The study used three methods to iteratively refine the dashboard: (1) user-centered design sessions with individuals who interact with the BEmOC supply chain, (2) review and feedback from domain and information visualization subject matter experts (SMEs) to refine the dashboard, and (3) usability heuristic evaluation with human-computer interaction (HCI) SMEs.

Results: User-centered design sessions resulted in a preliminary version of the dashboard informed by end-user preferences and perceptions, with recommendations focusing on aesthetic design, filtering and sorting, and matching with the real world. An example of an end-user recommendation included increasing font sizes on the dashboard and using a red, yellow, and green color-coding scheme. Next, domain and visualization SMEs continued the dashboard's iterative refinement, focusing on aesthetic design and navigation, by confirming design choices incorporated from the user-centered design sessions and recommending changes to enhance user experience moving through the dashboard, such as adding more filtering options. HCI SMEs rated the dashboard as highly usable (0.82 on a scale of 0-4, with 0 being no usability concern and 4 being a catastrophic usability concern). The principle with the highest usability severity scores was a match between the system and the real world with a score of 1.4. The HCI SMEs also rated the information visualization aspects of the dashboard favorably with 2 usability principles, spatial organization and information coding, scoring 0.

Conclusions: Dashboards are a novel method for promoting and tracking facility capacity to manage BEmOC. By including targeted end users and SMEs in the design process, the team was able to tailor the dashboard to meet user needs, fit it into the existing government health systems, and ensure that the dashboard follows design best practices. Collectively, the novel, customized

BEmOC dashboard can be used to track and improve facility-level readiness in Amhara, Ethiopia, and similar global BEmOC facilities.

(JMIR Hum Factors 2025;12:e64131) doi:10.2196/64131

KEYWORDS

health information technology; design and evaluation; Ethiopia; usability; nursing informatics; user-centered design; basic emergency obstetric care; obstetric; nurse; user-centered; design; maternal mortality; maternal; develop; sub-Saharan Africa; Africa; dashboard; tracking; emergency care

Introduction

Maternal mortality is a critical public health issue, particularly in low-resource settings like sub-Saharan Africa, which accounts for over three-fourths of global maternal deaths [1]. Despite progress in reducing adverse maternal outcomes in countries like Ethiopia, high maternal mortality ratios persist, largely due to gaps and stockouts of essential supplies for managing obstetric emergencies [2,3]. Inadequate supplies for basic emergency obstetric care (BEmOC) can lead to delayed or suboptimal care. A study on facility-level readiness to manage BEmOC in Amhara, Ethiopia, using the Clinical Cascades reported a 63.3% mean overall BEmOC readiness for managing the 6 most common emergencies, which underscores the need for improving facility-level BEmOC readiness in this region [4].

A promising approach to improve facility-level BEmOC readiness is the development and implementation of health information technology (HIT) to monitor critical supply availability. HIT involves the processing, storage, and exchange of health information in an electronic environment, and has been shown to improve health outcomes and strengthen health systems in low-resource settings [5,6]. Electronic dashboards are a form of HIT that can be used to track critical information, provide alerts, assess performance indicators, develop reports, and customize data views, which makes them a useful tool for monitoring inventory data and facility-level readiness [7,8]. For example, dashboards were used in Ethiopia to monitor the supply inventory of malaria medications and prevent stockouts from occurring, and a simple dashboard was merged into the existing Integrated Pharmaceutical Logistics System (IPLS) in the Amhara and Oromia regions of Ethiopia, resulting in improved facility-level readiness to treat tuberculosis by ensuring health care facilities had adequate supplies [7,8]. Beyond Ethiopia, HIT has been used to anticipate medical stockouts and improve maternal health outcomes in low-resource settings [9,10]. This evidence highlights the potential for HIT to improve facility-level BEmOC readiness in Amhara, ultimately contributing to improved maternal care quality and reduced maternal mortality ratios.

Previous studies, including those conducted in low-resource settings and Ethiopia, have demonstrated the benefits of HIT and dashboards in improving facility-level readiness [5-8]. Furthermore, the growing utilization of technology and cellular services across Africa further supports the feasibility of implementing successful electronic solutions, driven by the increasing adoption of HIT [11,12]. Finally, Ethiopia has recognized the value of HIT in strengthening health care systems and has been implementing electronic medical records in its health care facilities since 2013 [13].

However, while dashboards can be a powerful tool for improving health outcomes, their success depends on ensuring high usability, enabling end users to effectively accomplish their tasks. Usability is defined 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” [14]. Despite the potential of HIT, many implementations have failed due to insufficient consideration of end users’ needs and the environment in which the technology is applied [15,16]. Similarly, while dashboards show promise for improving facility-level BEmOC readiness, their success relies on development that incorporates input from targeted end users, accounts for the context of use, and adheres to usability principles. This study aims to develop and assess the usability of an electronic dashboard for monitoring facility-level BEmOC readiness.

Methods

Overview

To create the initial version of the dashboard, the team incorporated findings from previous qualitative interviews with individuals who worked in Ethiopia’s obstetric emergency supply chain, used information and designs from the preexisting dashboards in Ethiopia’s IPLS, and followed best practices found in the current literature for dashboard development [17-20]. The initial dashboard and its components were developed in Microsoft PowerPoint (refer to Multimedia Appendix 1).

To ensure adequate fit between the dashboard, the required tasks, the current system, the environment, and intended users, the research team conducted participatory user-centered design sessions, performed informal evaluations with domain and information visualization subject matter experts (SMEs), and completed a heuristic usability evaluation with human-computer interaction (HCI) SMEs.

Ethics Considerations

The research team obtained ethics approval from Columbia University’s institutional review board (IRB) (IRB-AAAU2006), Emory University’s IRB (MOD005-STUDY00005335), and the Amhara Public Health Institute (NoH/RfftTlDlo7l44). Before collecting data, the research team provided participants with an information sheet and gave them time to ask questions about the study. Once all questions had been answered, the team obtained verbal informed consent from all participants before

collecting data. All ethics review boards waived the requirement for written documentation of informed consent because the study procedures met the criteria for minimal risk to participants. All data was de-identified and participants data was linked to a random identification number. Only KD had access to the list of participant names linked to the ID numbers. None of the targeted end-users, domain experts, or subject matter experts received compensation for their participation in the study. Human-computer interaction experts participating in the heuristic evaluation received a $50 amazon gift card for their participation.

Sample Selection

User-Centered Participatory Design Sample Selection

Consistent with existing literature in the field of participatory design, the team estimated that a sample size of 15 would be sufficient to reach design saturation, a state where the design sessions reveal no major revisions to the designs reviewed [21,22]. The research team purposively recruited individuals with firsthand experience working within the BEmOC supply chain in Amhara, Ethiopia, at both the regional and local or facility levels. Participants were required to be at least 18 years old. Study participants were Ethiopian citizens and full-time employees in one aspect of Amhara's BEmOC supply chain, such as hospital supply managers, regional hub employees, and pharmacists.

Informal Expert Review Sample Selection

The domain SME was a research team member with over 6 years of experience conducting research in Amhara, Ethiopia; evaluating the supply chain; and measuring facility-level readiness to manage BEmOC. The information visualization SMEs were individuals from Columbia University School of Nursing's Visualization Design Studio, where individuals with varying levels of expertise can learn and gain advice on their current projects. The discussions at this studio were led by 2 PhD-level faculty, a nurse scientist with expertise in visualization design and evaluation, and a human-factors engineer. Key contributions also came from other faculty members at the university who have training in the areas of nursing research and biomedical informatics. All information visualization SMEs have published in the areas of information visualization, HCI, and user experience.

Heuristic Evaluation Sample Selection

The study sample included 5 HCI SMEs who were recruited through the team's professional networks. Previous research has found that 5-8 SMEs can identify over 80% of HIT usability violations [23,24]. HCI SMEs were eligible for inclusion if they had conducted research or published in the field of user interfaces or information visualizations and had not previously evaluated the dashboard.

Procedures

User-Centered Design Procedures

Staff from Emory-Ethiopia Partnership and Amhara Regional Health Bureau identified BEmOC supply chain stakeholders and reached out to determine if they were interested in participating in this study. A member of the research team contacted interested individuals via phone call, email, and WhatsApp (Meta) messaging to explain the study in detail, answer any questions, and set a time to meet in person with the participant to obtain verbal consent and perform the research activities. We used multiple communication channels to reach potential participants; in particular, WhatsApp has been found to be a useful tool for recruiting participants, specifically in global settings [25,26]. Design sessions were conducted in English and Amharic, and data collection tools were available in both languages since Amharic is the national language and Ethiopia's health care professionals receive clinical training in English [27,28].

Design sessions lasted approximately 60 minutes and occurred in small groups of 2-4 people within the health care facility or health bureau offices. Participants were given printed copies of the initial dashboard and asked to verbalize their thought processes as they explored the visualizations and information in the preliminary dashboard [29-32]. Design sessions were audio recorded and transcribed into English for analysis by a team member who is bilingual in Amharic and English. The individual conducting the analysis is a native Amharic speaker with over a decade of experience translating Amharic to English for health care research. Given their extensive expertise and the time constraints of the study, the team opted not to perform back translations. The dashboard development followed an iterative process, consisting of 4 design sessions. After each session, the research team incorporated recommended changes into the dashboard, which were then presented in the subsequent sessions [30].

Informal Expert Review Procedures

The team presented the Microsoft PowerPoint dashboard to a domain expert and a group of information visualization SMEs. Team member (KD) used Zoom (Zoom Video Communications) screen-sharing capabilities to walk the domain expert through the dashboard and their various features. The domain expert then reviewed the dashboard and provided recommendations for improvement.

KD presented the same dashboard to information visualization SMEs at 2 sessions of the Visualization Design Studio [33]. KD walked the SMEs through the dashboard, answered questions about their intended use, and elicited feedback and recommendations. Collectively, both rounds of feedback were used to optimize and further develop a medium-fidelity version of the dashboard for use in the heuristic usability evaluation.

Heuristic Evaluation Procedures

The team used Figma to develop a medium-fidelity prototype, which allowed the simulation of the dashboard's workflow [34]. Using Zoom Pro, a Health and Insurance Portability Accountability Act (HIPAA)–compliant videoconference platform, and its screen-sharing technology, KD asked the HCI SMEs to complete several tasks using the dashboard and encouraged them to explore other components of the dashboard (Multimedia Appendix 2). During their exploration, participants completed a heuristic evaluation checklist to identify and rank the severity of the usability violations and had the opportunity

to explain the violations they encountered [23,24,32]. They were also asked to describe what they were thinking, feeling, and seeing while completing the various tasks. The sessions were video recorded, and KD took field notes. At the end of the sessions, the HCI SMEs and KD discussed potential solutions for the identified usability violations. Participants were asked to complete tasks during the heuristic usability evaluation mentioned in Textbox 1.

Textbox 1. List of tasks participants were asked to complete during the heuristic usability evaluation.

Heuristic usability evaluation tasks

- Filter the main screen to only view primary hospitals.
- Determine which health care facilities are not ready to manage retained placenta.
- Determine if [hospital name] is ready to manage hypertensive emergencies. If not, identify which items are missing.
- Determine if [hospital name] is ready to manage maternal sepsis or infection. If not, identify which items are missing.
- Determine what emergencies [hospital name] is ready to manage.
- For prolonged labor at [hospital name] filter the supplies table to see the items that are at the emergency order point.
- For prolonged labor at [hospital name] filter the supplies to see the items that are overstocked.
- Determine how many health care facilities are at risk for not being able to manage retained placentas.

The Heuristic Evaluation Checklist used in this study was guided by Nielsen's 10 usability heuristics (ie, best practices; Table 1) [24,35] and includes a 5-item scale where experts rate each usability heuristic on a scale of 0 (not a problem) to 4 (catastrophic violation; Table 2) [23,32,36]. The research team modified the original checklist to fit the needs of a medium-fidelity dashboard (Multimedia Appendix 2). For example, questions for heuristics related to error prevention, help users with errors, and help and documentation were excluded since those tasks could not be simulated with the medium-fidelity prototype. In addition, 3 usability heuristics specific to information visualizations in the dashboards were included: spatial organization, information coding, and orientation [19].

Table 1. Usability heuristics and their definitions used in the heuristic usability evaluation.

Usability heuristic	Type	Definition
Visibility of system status	HIT[a]	The system should always keep users informed about what is going on through appropriate feedback in a reasonable time.
System and real-world match	HIT	The system should speak the user's language, with words, phrases, and concepts familiar to the user. The system should follow real-world conventions and ensure the dashboard fits within the existing workflow and technology system.
User control and freedom	HIT	Users should be free to select and sequence tasks and make their own decisions regarding the cost of exiting current work. Users should have clearly marked "emergency exit" to leave the unwanted state.
Consistency and standards	HIT	Users should not have to wonder whether different words, situations, or actions mean the same thing. Systems should maintain interface design choices in similar contexts and differ in different contexts.
Recognition rather than recall	HIT	The user should not have to remember information from one part of the dialogue to another. Objects, actions, and options should be easily visible, and instructions should be visible or easily retrievable whenever appropriate.
Flexibility and efficiency of use	HIT	The system should offer users several options for finding content. Users should be able to customize their interface and achieve their goals in an efficient manner and have the capacity to adapt to users' needs.
Aesthetic and minimalist design	HIT	The main dashboard should not contain information that is irrelevant or rarely needed. The system should present the largest amount of data with the least amount of ink.
Spatial organization	Information visualization	The overall layout of a visual representation should make it easy for the user to locate an information element in the display.
Information coding	Information visualization	The symbols and numbers used in the visualization should aid perception. The numeracy and graph literacy of the visualization should match the intended users' ability.
Orientation and help	Information visualization	The system should provide support for the user and help to orient them in their visualization.

[a]HIT: health information technology.

Table 2. Severity ratings for usability violations in the heuristic evaluation.

Score	Usability problem	Explanation
0	Not a usability problem at all	No problem identified
1	Cosmetic problem	Need not be fixed unless extra time is available on the project
2	Minor problem	Fixing this should be a low priority
3	Major usability	Important to fix, so should be given high priority
4	Usability catastrophe	Imperative to fix this before the product can be released

Data Analysis

User-Centered Design Analysis

Analysis was concurrent with data collection with modifications of the dashboard occurring after the initial 2 design sessions and following the third and fourth sessions. Using the translated transcripts of audio-recorded sessions and field notes, KD extracted key preferences, identified poorly performing graphics and dashboard components, and improved the design based on participant feedback. The team maintained detailed notes on design decisions and the rationale for those decisions alongside the notes and transcripts, which functioned as an audit trail.

Heuristic Usability Evaluation Analysis

The responses on the Heuristic Evaluation Checklist were transferred into Microsoft Excel for analysis [19,20]. The research team used unique participant numbers to link the responses to a confidential participant list. The data were analyzed with descriptive statistics (frequencies and means) [37]. Both session field notes and qualitative responses reported in the free response comments section of the checklist were used to understand usability and enhance the dashboard's design. During the evaluations, KD documented key areas of concern that HCI SMEs verbalized during the sessions. Combing those notes with the free-response comments, KD used the frequency that those concerns were mentioned to understand their importance. The usability concerns and potential solutions reported from the free-response comments were also compiled into a single list of action items, with duplicates removed, and KD used this list to guide the refinement of the next version of the dashboard.

Results

User-Centered Design Sessions

Sample Characteristics

In total, 6 individuals, who work at the regional health bureau or regional hubs for supply distribution, participated in the first 2 design sessions and reviewed the initial version of the dashboard. Following the refinement of the dashboard by KD, 5 individuals who work at health care facilities in the region participated in the third and fourth design sessions and reviewed the updated version of the dashboard. Recruitment and data collection ended prematurely at 11 participants due to civil unrest and safety concerns in the region [38].

Evolution of the Dashboard Design

During design sessions 1 and 2, participants reviewed the initial version of the prototype, and during design sessions 3 and 4, they reviewed a revised version that incorporated feedback from study participants in design sessions 1 and 2. During design sessions 1 and 2, participants viewed the initial design of the dashboard and were able to easily grasp the idea the dashboard was trying to portray related to facility readiness to manage BEmOC. The users felt the dashboard, specifically, the regional view, helped them to understand which facilities would be ready to handle the emergencies. One regional respondent stated,

> *This image [dashboard] demonstrates how they [individual health care facilities] would be able to handle each health issue considering the level of the supplies that each hospital has for these health problems. In general, we can conclude from this image [regional dashboard] that some hospitals are in good condition for these health issues and others are not in good condition for these health issues.*

Generally, participants in the first two sessions focused on (1) individual graphical component preferences such as which type of charts they preferred to see the data presented and (2) how they would want the data to be filtered and sorted. The third and fourth sessions with the refined dashboard focused on (1) confirming the terminology in the dashboard, and that it matched terms used in the supply chain, and (2) affirming design changes that were incorporated from the earlier sessions. Respondents in the third and fourth sessions also provided feedback on which job roles should have access to the dashboard. A participant during the fourth design session stated,

> *This dashboard [emergency-specific view] is essential for the dispensary unit as well. The store and supply officer bears most of the workload or duty in this workplace, but if the pharmacists and hospital managers have access to the data, it would be very good.*

The changes that were incorporated during the design sessions, as well as the session number, iterations of the dashboard, and job level that provided the recommendations, can be found in Multimedia Appendix 3. In the following paragraphs, more details are provided on aesthetic design, filtering and sorting, and matching with real-world preferences.

Aesthetic Design

For colors used throughout the dashboard, participants wanted to use multiple colors that were consistent with those currently used in the existing dashboard. Furthermore, participants in the

first 2 design sessions identified pieces of information that they thought should be removed from the dashboard because they did not assist the end users with making decisions related to supply ordering or maintaining facility-level readiness. However, to ensure users could find information that was not included in the dashboard, the research team added a button that would allow users to view the comprehensive drug list, which is already embedded in the electronic components of the existing IPLS. Facility-level participants, during design sessions 3 and 4, explained that not all health posts will manage certain obstetric emergencies, such as sepsis, so they should not receive readiness classifications for those emergencies. In response to this, the research team grayed out the classifications for those facilities and their corresponding emergencies.

Another aesthetic preference in the early design sessions included requests to include keys in the dashboard that define the colors and terms. Participants in design sessions 3 and 4 agreed that the keys assisted in their comprehension of the dashboard. Participants in design sessions 3 and 4 affirmed the design choices of the previous sessions and provided only small recommendations for the overall look of the dashboard, such as increasing the font size. Related to aesthetic decisions, design saturation was met with the smaller sample size.

Filtering and Sorting

Participants in the early sessions wanted to ensure there were multiple ways to filter the data so they could view pieces of information that were most important to them. Preferred filtering and sorting options included viewing inventory alphabetically within preestablished categories from the Ethiopian Pharmaceutical Supply System (eg, medical equipment, pharmaceuticals, and supplies), filtering supplies based on quantity available at the facility, and filtering facility by health care facility tier (eg, primary, general, referral, and center) [17]. Participants during the third and fourth design sessions endorsed the filtering options identified in the earlier sessions.

The participants from design sessions 3 and 4 also wanted employees at individual facilities to be able to view inventory data at other facilities besides their own. They felt having access to these data could support transferring excess supplies from one facility to another when neighboring facilities faced stockouts or were at an emergency order point. Thus, the research team added the ability for facility-level users to select different facility readiness data, similar to the regional-level users. Since there was very little new information identified during the second 2 design sessions, the research team determined that the study achieved design saturation for filtering preferences even with the smaller respondent sample.

Match With the Real-World Preferences

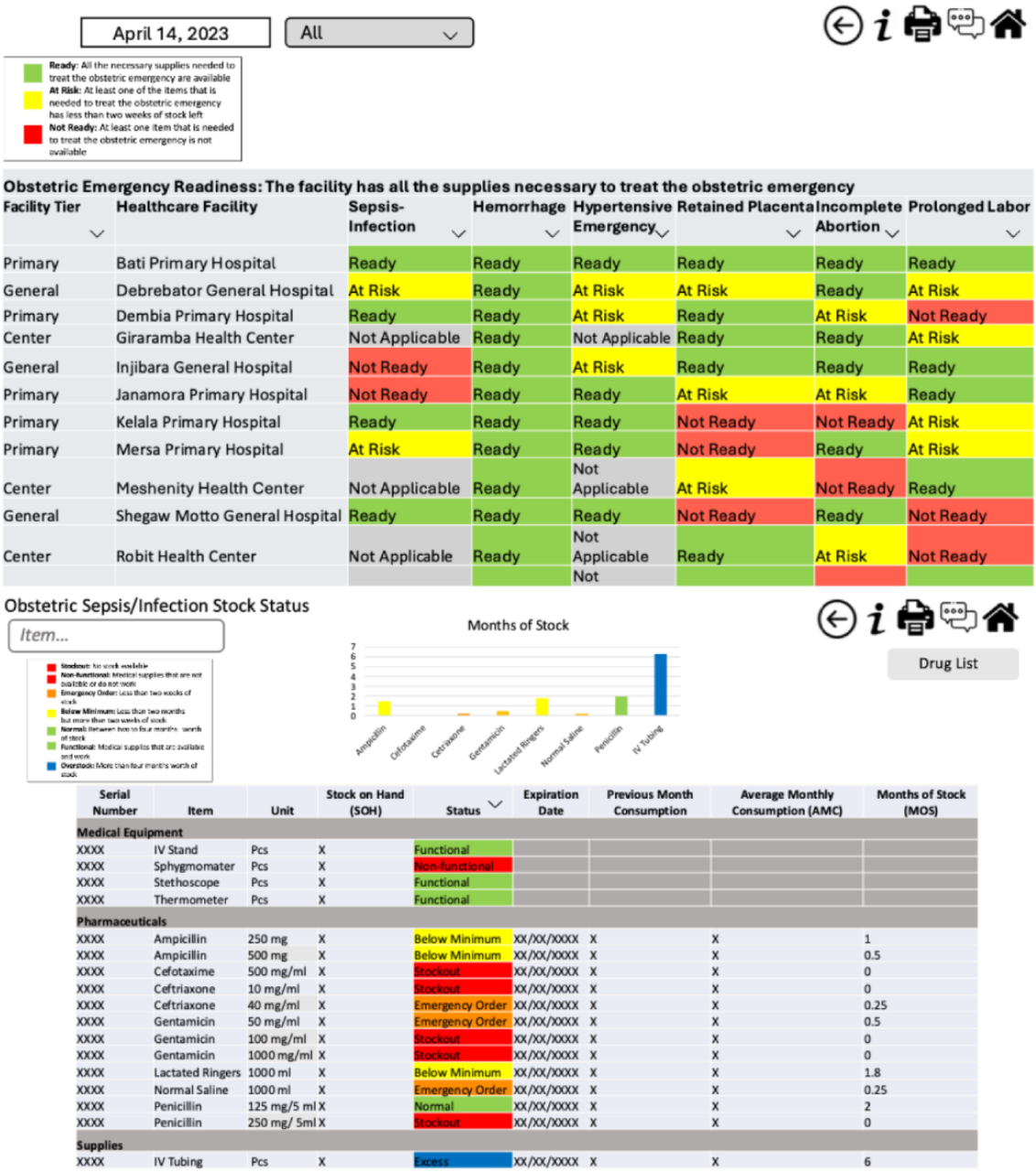
A feature that was a very important point of discussion in all the design sessions was using appropriate terminology that aligned with the participants' current workflow and job expectations. Participants wanted to ensure that all the terms used within the dashboard aligned with terms from the IPLS. For example, supply names and categories as well as terms for quantities available were updated (eg, overstock changed to excess, and normal changed to functional for the category medical supplies).

One important change to note is that during the third design session, participants recommended changing the terminology for BEmOC readiness from ready, at risk, and not ready to normal, emergency order, and stockout so they would be the same as the terms used with the emergency-specific dashboard view. Based on this recommendation, the research team created 3 different readiness terminology keys. The 3 options included the original terms; the newly recommended terms; and a third option of yes, no, and emergency order. Since the team could not confirm their preferred choice for readiness terminology due to the civil conflict in the region, the domain expert selected the final terms—yes, at risk, and no. Since the team was unable to obtain user preferences related to this final change, the study did not achieve design saturation in the "match with the real-world" domain.

Final Design

The final design comprised 2 different dashboard views for monitoring BEmOC readiness—one for the regional level and another for the facility level. The regional-level dashboard view provides a summary of readiness for the 6 obstetric emergencies for all health care facilities in the region. The facility-level dashboard view focuses on one facility and one obstetric emergency at a time. Overall study participants viewed the dashboard positively, with one respondent from design session 4 reporting "The hard copies [of the dashboard] you showed us today are beautiful and wonderful." Figure 1 shows the design of the dashboard following the integration of the findings from the third and fourth user-centered design sessions, and Multimedia Appendix 3 lists all the changes that were incorporated into the dashboard following the user-centered design sessions.

Figure 1. Regional basic emergency obstetric care (BEmOC) readiness dashboard and emergency-specific, facility-level dashboard following user-centered design sessions. Hospital names are real, but all readiness data are fictitious.



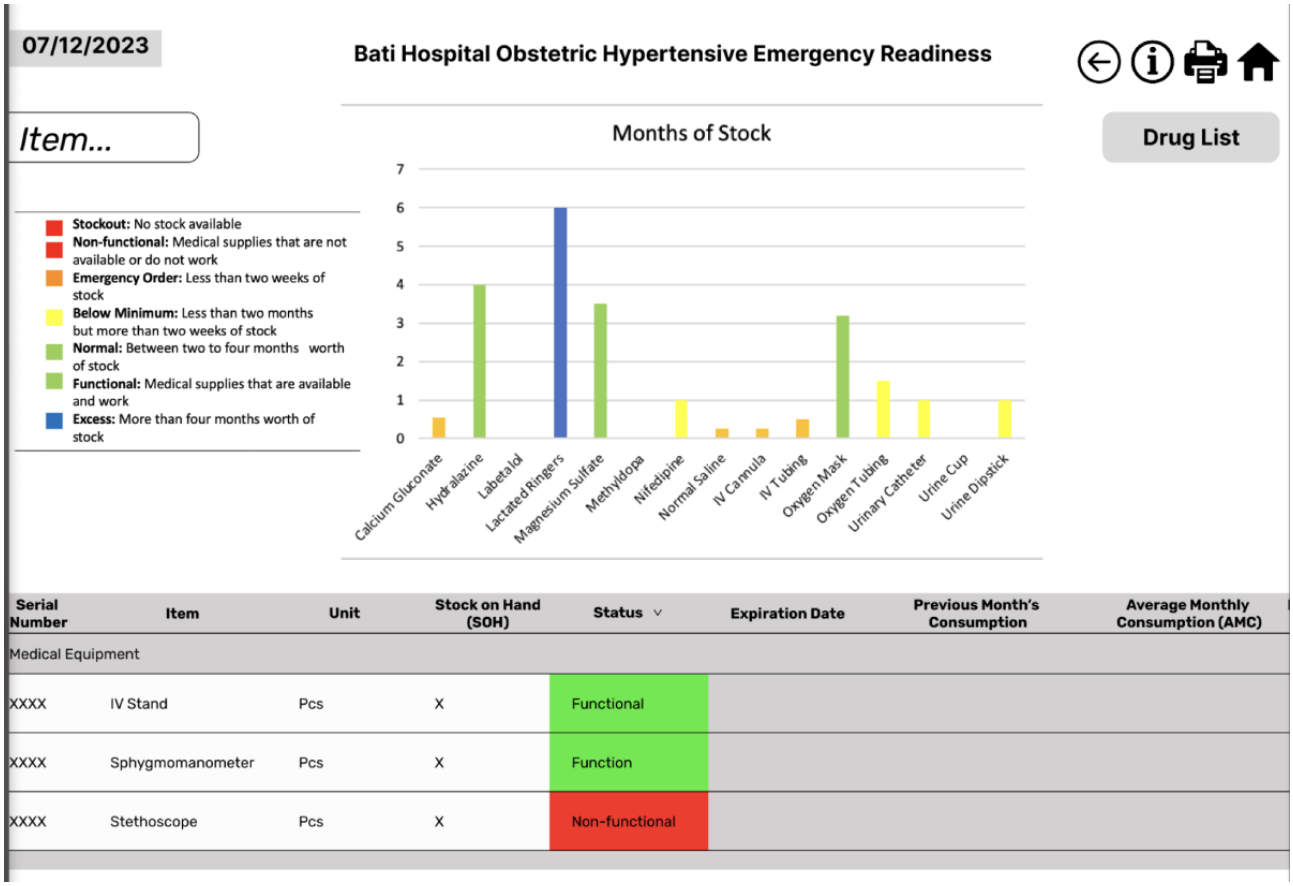
Informal Expert Review

The domain expert's largest contribution to the refinement of the dashboard was his input on which terms to use to define BEmOC readiness, specifically for the regional view dashboard. To make the final decision on readiness classification terms that were discussed in the final user-centered design session, KD presented the domain expert with 3 different options created following the fourth design session. After reviewing the options, the domain expert recommended KD use yes, emergency order, and no. The rationale for this decision was the yes and no choice was the simplest option, and he believed the end users would be able to readily act based on these classification terms. The domain expert also affirmed KD's decisions to group data at the regional and facility levels and felt that no critical information was missing from the dashboard.

During the review by information visualization SMEs, most recommendations aesthetics related to aesthetics and navigation. In terms of aesthetic design, the recommendations mainly focused on where to place different pieces of information or graphics to best grab the users' attention or prevent overcrowding on a screen, for example, moving several columns of data to the right in the emergency-specific dashboard view. This would allow users to scroll horizontally to see that information, but it would not be visible without scrolling so there would be less information crowding the page. To improve navigation between the various dashboard views, the SMEs during the design studio recommended adding additional filtering options, such as for emergency readiness categories. They also recommended adding an affordance feature to the dashboard so that users would be able to see which items on the screen are clickable or not. In addition, the experts

recommended adding several additional titles to different dashboard views to enhance clarity. These changes were incorporated when the dashboard was updated from PowerPoint to Figma (Figure 2).

Figure 2. Regional view dashboard following informal expert review. All inventory data in this dashboard are fictitious.



Heuristic Usability Evaluation With Human-Computer Interaction Experts

The heuristic evaluation took approximately 30 minutes per session. The years of experience for HCI SMEs ranged from 4 to 14 years, with a mean of 9. All HCI SMEs were female.

Overall, severity scores for heuristic evaluation ranged from 0 to 3 for the 10 usability principles. The mean usability score for all ten heuristics combined was 0.82 (SD 0.98), with a mode of zero. The principle with the highest severity rating was a match between the system and the real world, with a score of 1.4, which falls between a minor and cosmetic usability concern. For this principle, 4 HCI SMEs identified inappropriate color choices for both potential users with color blindness and using colors not congruent with common color-coding expectations as a common issue. Additional problems included the dashboard using concepts and phrases that were unfamiliar to the evaluators, and 3 evaluators felt the section headings were not ordered in the most logical fashion. HCI SMEs identified issues with usability principles visibility of system status, user control and freedom, consistency and standards, and recognition rather than recall (mean 1.2 for each, SD 0.84, 1.3, 0.84, and 1.3, respectively), which also falls between a cosmetic and minor usability concern. Concerns related to these principles included 2 of the HCI SMEs identifying the lack of clear exits for every dashboard and feeling that the current design hindered their ability to efficiently navigate the system. Table 3 shows the mean value, range, and mode for the 10 heuristic principles.

Table 3. Identified heuristic usability problems and their severity scores.

Heuristic category	Problems	Score, mean (SD)	Score, range	Score, mode
Visibility of system status	• Title font is too small • Color choices may be inappropriate for color-blind users • Vertical bar charts make reading supply labels more difficult • Need hyperlinks to improve navigation • Menu-naming terminology is not consistent with the user's task domain	1.2 (0.84)	2	1 and 2
Match between system and real-world	• Color choices may be inappropriate for color-blind users • Section headings and subheadings are not ordered in the most logical way • Not all words, concepts, and phrases were familiar to the human-computer interaction subject matter experts • Some of the colors used in the dashboard do not correspond to common expectations about color codes	1.4 (0.89)	2	1
User control and freedom	• Need hyperlinks to improve navigation • Not a clear exit on each dashboard screen • Not all screens are accessible across the system • Users could not easily move forward and backward between fields	1.2 (1.3)	3	0
Consistency and standards	• Titles on the regional view dashboard do not update when the user filters the data • Abbreviations not clearly explained • Some colors are too similar to distinguish from other categories • Not enough or inconsistent visual cues to identify active screens	1.2 (0.84)	2	1 and 2
Recognition rather than recall	• White space is not optimized within the emergency-specific dashboard view • Prompts, cues, and messages are not placed where the eye is likely to be looking on the screen	1.2 (1.3)	3	0
Flexibility and efficiency of use	• Need hyperlinks to improve navigation	0.6 (0.89)	2	0
Aesthetic and minimalist design	• Not all field labels are brief, familiar, or descriptive • Large objects, bold fonts, and simple areas have not been used to distinguish sections • There is not enough white space between color representation • Too much text is present in the keys which makes the screens look busy	0.8 (0.84)	2	0 and 1
Spatial organization	• Font size is too small throughout the dashboard • Information does not follow a logical flow	0 (0.0)	0	0
Information coding	• No problems identified	0 (0.0)	0	0
Orientation	• Measurement units are not displayed clearly • Users cannot control the level of detail they see in a representation	0.6 (1.3)	3	0

The experts provided several recommendations for how to remedy the problems identified during the heuristic evaluation. They believed adding hyperlink functions would improve users' ability to navigate between the dashboard views, thus increasing usability scores for visibility of system status, user control and freedom, and flexibility and efficiency of use. Aesthetically, the experts believed increasing the font size of titles and data within the dashboard, as well as moving different components of the dashboard around to maximize white space would improve scores for spatial organization, aesthetic and minimalist design, recognition rather than recall, and visibility of system status.

HCI SMEs also made recommendations that disagreed with user preferences identified during the design sessions. These differences included the experts preferring horizontal bar charts compared to vertical bar charts. An expert explained the rationale for this preference stating a horizontal bar chart allows for the labels in the table to be read horizontally, which can be easier for users. However, during the user-centered design sessions, the participants had a resounding preference for the vertical bar charts because they were more familiar with that format since some of their preexisting dashboards included vertical bar charts in the IPLS. Given the importance of the new dashboard to mimic the look of the existing ones, as well as the

fact that the users explicitly preferred the vertical presentation, KD chose to retain the vertical bar charts. The color of the tables and graphs was also a point of disagreement between experts and users. The experts did not think the red, yellow, and green color choices were appropriate because these colors can be difficult to distinguish if individuals are color blind [39]. However, the targeted end users of the dashboard preferred those colors because the colors reminded them of traffic lights, and the users could make assumptions related to the significance of the colors based on what those colors mean by a traffic light. The research team chose to keep the traffic light color scheme.

Beyond these recommendations, the reviewers cited several positive features of the dashboard. This includes things such as the belief that the icons used throughout the dashboard were clear and easy to associate with their function. In addition, the experts found the labels and keys to be clear and assisted with comprehension of the data. One respondent reported, "Your labels are quite clear. It's [the dashboard] very accessible and you have nice keys right here." The heuristic violation severity scores for information visualization-specific heuristics were low with scores of zero for spatial organization and information coding, and 0.6 for orientation. Table 4 summarizes the changes incorporated into the dashboard following the completion of the heuristic usability evaluation. The final dashboard is displayed in Figure 3.

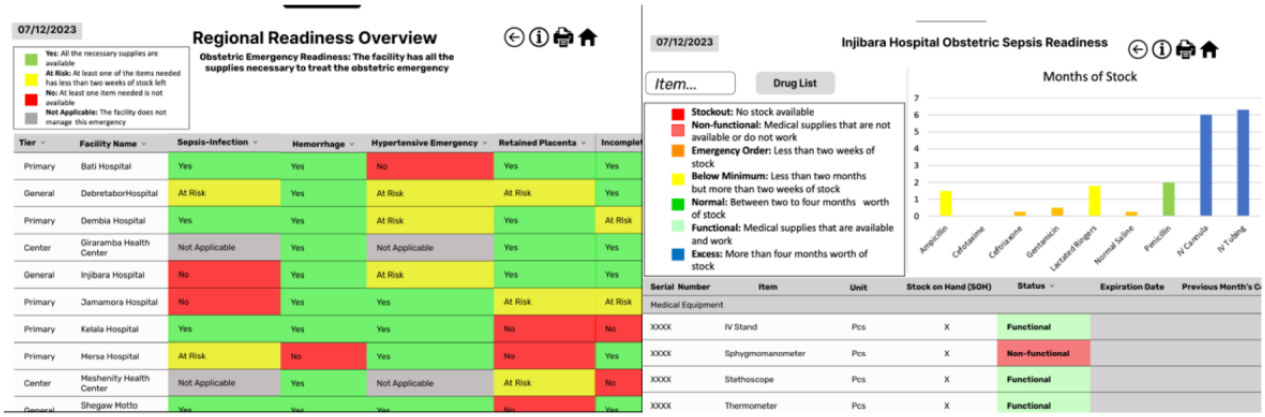
Table 4. Summary of changes incorporated into the dashboard following the heuristic evaluation

Heuristic category	Changes following heuristic evaluation
Visibility of system status	• Increased font size throughout the dashboard • Added hyperlinks to the dashboard
Match between system and real-world	• Updated section headings to reflect filtering capabilities • Updated the terminology defining BEmOCa readiness on the regional-level dashboard
User control and freedom	• Added hyperlinks to the dashboard • Ensured exit buttons on every screen are activated • Activated more functions within the dashboard to make it a higher-fidelity prototype
Consistency and standards	• Updated section headings to reflect filtering capabilities • Defined all abbreviations used in the dashboard • Changed the colors for functional and nonfunctional on the emergency-specific screen so that they are more easily distinguishable from the other colors used on the screen • Added additional visual cues to identify active screens
Recognition rather than recall	• Moved graphics at the top of the dashboard to optimize white space • Added additional visual cues to identify active screens
Flexibility and efficiency of use	• Added hyperlinks to the dashboard
Aesthetic and minimalist design	• Shortened information in the keys to make them one line of text • Shortened data within the tables to make it one line of text • Used bold fonts for all titles • Moved graphics at the top of the dashboard to optimize white space
Spatial organization	• Increased font size throughout the dashboard • Improved the navigation between dashboard views to make them a higher-fidelity prototype
Information coding	• N/Ab
Orientation	• Increased font size for measurement units • Activated more functions within the dashboard to make it a higher-fidelity prototype

[a]BEmOC: basic emergency obstetric care.

[b]N/A: not applicable.

Figure 3. Final regional view and emergency-specific view of the dashboard following the heuristic usability evaluation. Hospital names are real, but all readiness data are fictitious.



Discussion

Principal Findings

The dashboard underwent iterative refinement based on user input, domain expert feedback, information visualization evaluations, and heuristic usability assessments. Regional participants emphasized preferences for vertical bar charts, traffic-light color coding, and keys to define readiness levels, while facility-level users focused on aligning terminology with existing workflows and tailoring readiness classifications. The information domain expert provided critical input on readiness terminology, and information visualization experts suggested improvements in layout, navigation, and aesthetics, such as optimizing white space and increasing font sizes. Collectively, the central tendency of heuristic severity scores, combined with a mode of zero, indicated high overall usability of the dashboard. However, the HCI experts did highlight minor usability concerns, including navigation and accessibility improvements, which were addressed in the final design. While some expert recommendations, such as horizontal bar charts and alternative color schemes, differed from user preferences, the team prioritized end-user input to ensure practical usability. The final dashboard is intuitive, visually clear, and aligned with real-world needs, achieving high usability and user satisfaction.

This study highlights the importance of including participants from all targeted end-user groups. By including employees from both regional and facility levels, the team gathered diverse, relevant perspectives, ensuring the dashboard met the real-world needs of users across different supply chain roles and experiences. Feedback from both end users and experts enhanced the overall usability and functionality of the dashboard. This aligns with existing usability literature, which emphasizes the value of involving both end users and experts to identify distinct concerns and improve HIT design [40]. Participants' spontaneous interpretations of the dashboard underscore their understanding of the content, and obtaining these perceptions and insights can assist in preventing usability errors from occurring later in the piloting or implementation phase of HIT development. These refinements will likely maximize the ease of implementation and promote consistent uptake and utilization by future users [41,42].

There were 2 areas of disagreement between HCI SMEs' recommendations and end-user preferences: bar chart orientation and color schemes. The research team chose to keep the vertical bar charts for their familiarity with the end users and alignment with existing dashboards. In addition, lower color blindness prevalence among Black African and African American populations (4.2% compared to 6.6% and 1.4% compared to 5.6% respectively) compared to White populations (6.6% and 5.6%) [43,44] supported the decision to maintain the preferred color scheme, as most targeted end users were unlikely to face difficulties distinguishing the colors.

This study reinforces the notion that obtaining expert opinions and performing usability testing can be a low-cost, efficient method for exploring technology needs [45]. Performing the evaluations with the domain expert and the information visualization SMEs took less than an hour for both groups and provided critical information on how to improve upon the design of the dashboard. Performing usability evaluations can identify problems and usability concerns that may not have been noticed before HIT implementation, avoiding costly and time-consuming corrections [24].

Finally, this study contributes to the growing body of literature on the use of dashboards to support BEmOC in sub-Saharan Africa. Similar to findings by Banke-Thomas et al [46] and Wang et al [47], stakeholders expressed enthusiasm for dashboards and prioritized features like reporting facility characteristics on the dashboards. However, there remains a gap in the literature regarding the specific design preferences of end users. This study addresses that gap by highlighting key preferences, such as the use of stoplight color coding and IPLS-specific terminology, which enhance usability and are tailored specifically for dashboards intended for Ethiopia and BEmOC contexts.

Limitations

During data collection, there was unanticipated civil unrest in the Amhara region, which prevented the research team from recruiting participants outside of the regional capital of Bahir Dar and from completing participant recruitment from multiple facilities outside the regional capital. Despite this unforeseen barrier, design saturation was reached for 2 out of the 3 categories during the design sessions. There was also a lack of

gender diversity among our participants. Only one user-centered design participant from Ethiopia was female. However, this gender imbalance mirrored the existing gender imbalance in the existing health system for these roles. However, future research related to this dashboard and facility-level BEmOC readiness would benefit from an emphasis on opinions and participation from female participants with relevant expertise. Conversely, all HCI SMEs were female. While the incorporation of a female perspective is critical since the sample for user-centered design sessions was predominately male, the effect of having completely female HCI SMEs is unknown. Future research should incorporate technical and gender-balanced perspectives during both user and HCI SMEs' review of this dashboard.

Furthermore, the dashboard design was informed by responses from participants in qualitative interviews. However, due to time constraints, the initial prototype was developed by the research team rather than being cocreated with participants. While this approach allowed the team to efficiently move forward with development, it may have limited the incorporation of deeper, iterative feedback from participants during the initial design phase. This limitation could have influenced the alignment of the prototype with user needs and preferences, which may require further refinement in future iterations.

To prospectively limit social desirability bias, the research team created a low-stress, comfortable environment for data collection and emphasized our desire to obtain personal responses to the dashboard while reiterating that all opinions would be blinded so no one could identify specific feedback from specific participants. The team chose to have an Amharic-speaking Ethiopian team member conduct the user-centered design sessions rather than the principal investigator, who is from the United States and does not speak Amharic. The team took steps to ensure confidentiality and used indirect questioning techniques in an attempt to garner the participants' true opinions [48]. Furthermore, the dashboard reviewed in the heuristic usability evaluation was medium fidelity. This means the dashboard was not able to complete all required tasks and certain usability heuristics, such as *help and documentation* and *error prevention*, were excluded from this evaluation, since their tasks were not applicable at this stage in the design process. Future research will need to explore all usability heuristics before implementation to ensure no new concerns arise once the dashboard transitions from a medium- to high-fidelity model.

Conclusion

In conclusion, this study demonstrates the value of user-centered design and usability evaluations in developing HIT for low-resource settings. By integrating diverse perspectives from regional and facility-level participants, as well as domain and usability experts, the dashboard was refined to meet real-world needs and achieve high usability. This iterative approach not only addressed usability concerns but also incorporated end-user preferences, ensuring alignment with existing workflows. The study fills a gap in the literature by identifying design preferences tailored to Ethiopia's health care context, offering a model for future HIT development aimed at improving BEmOC readiness and maternal health outcomes. These findings emphasize the importance of engaging end users and experts early in the design process to create functional, user-friendly systems that support effective implementation and long-term adoption.

Acknowledgments

We would like to thank Emory-Ethiopia Partnership's Addis Ababa and Bahir Dar offices for their support and guidance in this research.

This work was supported by the National Institute of Nursing Research (F31NR020569), a Fulbright US Student research award to Ethiopia, and a research grant from Sigma Theta Tau Alpha Zeta chapter.

Authors' Contributions

YA performed data collection with support from MB and LA. HB supported the project administration and maintained ethics approvals throughout the study. KD analyzed the data and created the dashboard with assistance from SB and JNC. KD drafted the manuscript, and all other authors assisted in editing and refining the final manuscript. JNC conceptualized the nested study design and obtained parent funding for the Saving Little Lives study which this research was nested within. AG provided contextual expertise on Amhara and Ethiopian clinical care.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The graph options and preliminary dashboard prototypes used in the user-centered design sessions.
[DOCX File , 206 KB - humanfactors_v12i1e64131_app1.docx]

Multimedia Appendix 2
Heuristic usability evaluation survey.
[DOCX File , 28 KB - humanfactors_v12i1e64131_app2.docx]

Multimedia Appendix 3
Changes incorporated into the dashboard following the user-centered design sessions.
[DOCX File , 22 KB - humanfactors_v12i1e64131_app3.docx]

References

1. Trends in maternal mortality 2000 to 2017: estimates by WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division. World Health Organization. 2021. URL: https://www.unfpa.org/featured-publication/trends-maternal-mortality-2000-2017 [accessed 2025-02-20]
2. Emory-Ethiopia partnership. Emory University's Nell Hodgson Woodruff School of Nursing. URL: https://www.nursing.emory.edu/overview/emory-ethiopia-partnership [accessed 2021-07-20]
3. Tessema GA, Laurence CO, Melaku YA, Misganaw A, Woldie SA, Hiruye A, et al. Trends and causes of maternal mortality in Ethiopia during 1990-2013: findings from the Global Burden of Diseases Study 2013. BMC Public Health 2017;17(1):160 [FREE Full text] [doi: 10.1186/s12889-017-4071-8] [Medline: 28152987]
4. Dougherty K, Gebremariam Gobezayehu A, Lijalem M, Alamineh Endalamaw L, Biza H, Cranmer JN. Comparison of obstetric emergency clinical readiness: a cross-sectional analysis of hospitals in Amhara, Ethiopia. PLoS One 2023;18(8):e0289496 [FREE Full text] [doi: 10.1371/journal.pone.0289496] [Medline: 37535678]
5. Brenner SK, Kaushal R, Grinspan Z, Joyce C, Kim I, Allard RJ, et al. Effects of health information technology on patient outcomes: a systematic review. J Am Med Inform Assoc 2016;23(5):1016-1036 [FREE Full text] [doi: 10.1093/jamia/ocv138] [Medline: 26568607]
6. Ngwa W, Olver I, Schmeler KM. The use of health-related technology to reduce the gap between developed and undeveloped regions around the globe. Am Soc Clin Oncol Educ Book 2020;40:1-10 [FREE Full text] [doi: 10.1200/EDBK_288613] [Medline: 32223667]
7. Fact sheet Ethiopia. USAID. 2020. URL: https://www.usaid.gov/sites/default/files/2022-05/Ethiopia-Fact-Sheet_PMI_July-2020.pdf [accessed 2025-02-20]
8. Rabiei R, Almasi S. Requirements and challenges of hospital dashboards: a systematic literature review. BMC Med Inform Decis Mak 2022;22(1):287 [FREE Full text] [doi: 10.1186/s12911-022-02037-8] [Medline: 36348339]
9. Jbaily A, Feldhaus I, Bigelow B, Kamareddine L, Tolla MT, Bouvier M, et al. Toward health system strengthening in low- and middle-income countries: insights from mathematical modeling of drug supply chains. BMC Health Serv Res 2020;20(1):776 [FREE Full text] [doi: 10.1186/s12913-020-05549-z] [Medline: 32838778]
10. Till S, Mkhize M, Farao J, Shandu LD, Muthelo L, Coleman TL, et al. Digital health technologies for maternal and child health in Africa and other low- and middle-income countries: cross-disciplinary scoping review with stakeholder consultation. J Med Internet Res 2023;25:e42161 [FREE Full text] [doi: 10.2196/42161] [Medline: 37027199]
11. Bukachi F, Pakenham-Walsh N. Information technology for health in developing countries. Chest 2007;132(5):1624-1630. [doi: 10.1378/chest.07-1760] [Medline: 17998362]
12. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the center for health market innovations. Bull World Health Organ 2012;90(5):332-340 [FREE Full text] [doi: 10.2471/BLT.11.099820] [Medline: 22589566]
13. Bekele TA, Gezie LD, Willems H, Metzger J, Abere B, Seyoum B, et al. Barriers and facilitators of the electronic medical record adoption among healthcare providers in Addis Ababa, Ethiopia. Digit Health 2024;10:20552076241301946 [FREE Full text] [doi: 10.1177/20552076241301946] [Medline: 39659403]
14. Usability of consumer products and products for public use-part 2:summative test method. International Organization for Standardization. URL: https://www.iso.org/obp/ui/#iso:std:iso:ts:20282:-2:ed-2:v1:en [accessed 2021-10-09]
15. Abugabah A, Alfarraj A. Issues to consider in designing health care information systems: a user-centred design approach. Electronic Journal of Health Informatics 2015;9.
16. Gnanlet A, Choi M, Davoudpour S. Impediments to the implementation of healthcare information technology : a systematic literature review. Journal of Supply Chain and Operations Management 2019;17(1).
17. Standard operating procedures (SOP) manual for the Integrated Pharmaceuticals Logistics System in health facilities of Ethiopia. International Institute for Primary Health Care - Ethiopia. 2020. URL: http://repository.iphce.org/handle/123456789/2613 [accessed 2025-02-20]
18. Dougherty K, Gebremariam A, Biza H, Belew M, Benda N, Tesfaye Y, et al. Obstetric emergency supply chain dynamics and information flow among obstetric emergency supply chain employees: key informant interview study. JMIR Form Res 2024;8:e59690 [FREE Full text] [doi: 10.2196/59690] [Medline: 39235860]
19. Dowding D, Merrill JA. The development of heuristics for evaluation of dashboard visualizations. Appl Clin Inform 2018;9(3):511-518 [FREE Full text] [doi: 10.1055/s-0038-1666842] [Medline: 29998455]
20. Nielsen J. 10 Usability heuristics for user interface design. Nielsen Norman Group. 1994 Apr 24. URL: https://www.nngroup.com/articles/ten-usability-heuristics/ [accessed 2025-02-20]
21. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]

22. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods 2006;18(1):59-82. [doi: 10.1177/1525822X05279903]
23. Allen M, Currie LM, Bakken S, Patel VL, Cimino JJ. Heuristic evaluation of paper-based Web pages: a simplified inspection usability methodology. J Biomed Inform 2006;39(4):412-423 [FREE Full text] [doi: 10.1016/j.jbi.2005.10.004] [Medline: 16321575]
24. Nielsen J. Finding usability problems through heuristic evaluation. 1992 Presented at: CHI '92: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 3-7, 1992; Monterey, CA p. 373-380 URL: https://doi.org/10.1145/142750.142834 [doi: 10.1145/142750.142834]
25. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022;24(8):e38015 [FREE Full text] [doi: 10.2196/38015] [Medline: 35925655]
26. Manji K, Hanefeld J, Vearey J, Walls H, de Gruchy T. Using WhatsApp messenger for health systems research: a scoping review of available literature. Health Policy Plan 2021;36(5):594-605 [FREE Full text] [doi: 10.1093/heapol/czab024] [Medline: 33860314]
27. Medicine. Addis Ababa University. 2021. URL: http://www.aau.edu.et/medicine/ [accessed 2025-02-20]
28. Nursing. Addis Ababa University. 2021. URL: http://www.aau.edu.et/nursing/ [accessed 2025-02-20]
29. Lai TY, Larson EL, Rockoff ML, Bakken S. User acceptance of HIV TIDES--tailored interventions for management of depressive symptoms in persons living with HIV/AIDS. J Am Med Inform Assoc 2008;15(2):217-226 [FREE Full text] [doi: 10.1197/jamia.M2481] [Medline: 18096915]
30. Lucero RJ, Sheehan B, Yen PY, Velez O, Nobile-Hernandez DL, Tiase VL, et al. Developing self-management tools with vulnerable populations for use in personal health information management systems. NI 2012 (2012) 2012;2012:248 [FREE Full text] [Medline: 24199096]
31. Vélez O, Okyere PB, Kanter AS, Bakken S. A usability study of a mobile health application for rural Ghanaian midwives. J Midwifery Womens Health 2014;59(2):184-191 [FREE Full text] [doi: 10.1111/jmwh.12071] [Medline: 24400748]
32. Yen PY, Bakken S. A comparison of usability evaluation methods: heuristic evaluation versus end-user think-aloud protocol - an example from a web-based communication tool for nurse scheduling. AMIA Annu Symp Proc 2009;2009:714-718 [FREE Full text] [Medline: 20351946]
33. Arcia A, Benda N. Visualization Design Studio. URL: https://vis-studio.mailchimpsites.com/ [accessed 2023-08-08]
34. Figma. 2016. URL: https://www.figma.com/ [accessed 2025-02-20]
35. Nielsen J, Molich R. Heuristic evaluation of user interfaces. 1990 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 1-5, 1990; Seattle, WA p. 249-256 URL: https://doi.org/10.1145/97243.97281 [doi: 10.1145/97243.97281]
36. Lai TY, Bakken S. Heuristic evaluation of HIV-TIDES - tailored interventions for management of depressive symptoms in HIV-infected individuals. AMIA Annu Symp Proc 2006;2006:996 [FREE Full text] [Medline: 17238615]
37. Choi J, Bakken S. Web-based education for low-literate parents in neonatal intensive care unit: development of a website and heuristic evaluation and usability testing. Int J Med Inform 2010;79(8):565-575 [FREE Full text] [doi: 10.1016/j.ijmedinf.2010.05.001] [Medline: 20617546]
38. Alene GD, Ayenew GM, Bantie GM, Zeru T, Yismaw G. Impact of ware on maternal health services in the war affected areas of the Amhara region, ethiopia: disruption of antenatla care, institutional delivery and postnatal services. Ethiopian Journal of Translational Sciences 2023;1(1).
39. Types of color vision deficiency. National Eye Institute. URL: https://www.nei.nih.gov/learn-about-eye-health/eye-conditions-and-diseases/color-blindness/types-color-vision-deficiency#:~:text=The%20most%20common%20type%20of,of%20green%20look%20more%20red [accessed 2025-02-20]
40. Følstad A. Users' design feedback in usability evaluation: a literature review. Hum Cent Comput Inf Sci 2017;7(1). [doi: 10.1186/s13673-017-0100-y]
41. Ergonomics of human-system interaction part 210: human-centred design for interactive systems. International Organization for Standardization. 2010. URL: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/20/52075.html [accessed 2023-08-01]
42. Lau MK, Bounthavong M, Kay CL, Harvey MA, Christopher MLD. Clinical dashboard development and use for academic detailing in the U.S. department of veterans affairs. J Am Pharm Assoc (2003) 2019;59(2S):S96-S103.e3. [doi: 10.1016/j.japh.2018.12.006] [Medline: 30713078]
43. Mashige KP, van Staden DB. Prevalence of congenital colour vision deficiency among Black school children in Durban, South Africa. BMC Res Notes 2019;12(1):324 [FREE Full text] [doi: 10.1186/s13104-019-4374-1] [Medline: 31182158]
44. Color blindness most common in White boys. Review of Optometry. URL: https://tinyurl.com/2ztwwm6w [accessed 2025-02-20]
45. Farzandipour M, Nabovati E, Sadeqi Jabali M. Comparison of usability evaluation methods for a health information system: heuristic evaluation versus cognitive walkthrough method. BMC Med Inform Decis Mak 2022;22(1):157 [FREE Full text] [doi: 10.1186/s12911-022-01905-7] [Medline: 35717183]

46. Banke-Thomas A, Abejirinde IO, Ogunyemi O, Gwacham-Anisiobi U. Innovative dashboard for optimising emergency obstetric care geographical accessibility in Nigeria: qualitative study with technocrats. Health Policy and Technology 2023;12(2):100756 [FREE Full text] [doi: 10.1016/j.hlpt.2023.100756]
47. Wang J, Wong KLM, Olubodun T, Gwacham-Anisiobi U, Ogunyemi O, Afolabi BB, et al. Developing policy-ready digital dashboards of geospatial access to emergency obstetric care: a survey of policymakers and researchers in sub-Saharan Africa. Health Technol 2023;14(1):69-80. [doi: 10.1007/s12553-023-00793-9]
48. Cobo B, Castillo E, López-Torrecillas F, Rueda MDM. Indirect questioning methods for sensitive survey questions: modelling criminal behaviours among a prison population. PLoS One 2021;16(1):e0245550 [FREE Full text] [doi: 10.1371/journal.pone.0245550] [Medline: 33481882]

Abbreviations

BEmOC: basic emergency obstetric care
HCI: human-computer interaction
HIPAA: Health and Insurance Portability Accountability Act
HIT: health information technology
IPLS: Integrated Pharmaceutical Logistics System
IRB: institutional review board
SME: subject matter expert

Edited by A Kushniruk; submitted 16.07.24; peer-reviewed by M Gasmi , A Banke-Thomas; comments to author 20.12.24; revised version received 21.01.25; accepted 10.02.25; published 03.04.25.

Please cite as:
Dougherty K, Tesfaye Y, Biza H, Belew M, Benda N, Gebremariam Gobezayehu A, Cranmer J, Bakken S
User-Centered Design of an Electronic Dashboard for Monitoring Facility-Level Basic Emergency Obstetric Care Readiness in Amhara, Ethiopia: Mixed Methods Study
JMIR Hum Factors 2025;12:e64131
URL: https://humanfactors.jmir.org/2025/1/e64131
doi:10.2196/64131
PMID:

Original Paper

Co-Designing a Web-Based and Tablet App to Evaluate Clinical Outcomes of Early Psychosis Service Users in a Learning Health Care Network: User-Centered Design Workshop and Pilot Study

Kathleen E Burch[1], BA; Valerie L Tryon[1], PhD; Katherine M Pierce[1], PhD; Laura M Tully[1], PhD; Sabrina Ereshefsky[1], PhD; Mark Savill[1,2], PhD; Leigh Smith[3], PhD; Adam B Wilcox[4], PhD; Christopher Komei Hakusui[1], BA; Viviana E Padilla[1], BA; Amanda P McNamara[5], MPH; Merissa Kado-Walton[5], MPH; Andrew J Padovani[6], PhD; Chelyah Miller[1], BA; Madison J Miles[1], BA; Nitasha Sharma[1], BA; Khanh Linh H Nguyen[1], MS; Yi Zhang[6], PhD; Tara A Niendam[1], PhD

[1]Department of Psychiatry & Behavioral Sciences, University of California-Davis, Sacramento, CA, United States

[2]Department of Psychiatry and Behavioral Sciences, Weill Institute for Neuroscience, University of California-San Francisco, San Francisco, CA, United States

[3]Department of Psychology, University of California-Davis, Davis, CA, United States

[4]Institute for Informatics, Washington University in St. Louis, St. Louis, MO, United States

[5]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, San Diego, CA, United States

[6]Center for Healthcare Policy and Research, University of California-Davis, Sacramento, CA, United States

Corresponding Author:
Kathleen E Burch, BA
Department of Psychiatry & Behavioral Sciences
University of California-Davis
4701 X St
Sacramento, CA, 95817
United States
Phone: 1 916 699 5193
Email: knye@ucdavis.edu

Abstract

Background: The Early Psychosis Intervention Network of California project, a learning health care network of California early psychosis intervention (EPI) programs, prioritized incorporation of community partner feedback while designing its eHealth app, Beehive. Though eHealth apps can support learning health care network data collection aims, low user acceptance or adoption can pose barriers to successful implementation. Adopting user-centered design (UCD) approaches, such as incorporation of user feedback, prototyping, iterative design, and continuous evaluation, can mitigate these potential barriers.

Objective: We aimed to use UCD during development of a data collection and data visualization web-based and tablet app, Beehive, to promote engagement with Beehive as part of standard EPI care across a diverse user-base.

Methods: Our UCD approach included incorporation of user feedback, prototyping, iterative design, and continuous evaluation. This started with user journey mapping to create storyboards, which were then presented in UCD workshops with service users, their support persons, and EPI providers. We incorporated feedback from these workshops into the alpha version of Beehive, which was also presented in a UCD workshop. Feedback was again incorporated into the beta version of Beehive. We provided Beehive training to 4 EPI programs who then piloted Beehive's beta version. During piloting, service users, their support persons, and EPI program providers completed Beehive surveys at enrollment and every 6 months after treatment initiation. To examine preliminary user acceptance and adoption during the piloting phase, we assessed rates of participant enrollment and survey completion, with a particular focus on completion of a prioritized survey: the Modified Colorado Symptom Index.

Results: UCD workshop feedback resulted in the creation of new workflows and interface changes in Beehive to improve the user experience. During piloting, 48 service users, 42 support persons, and 72 EPI program providers enrolled in Beehive. Data were available for 88% (n=42) of service users, including self-reported data for 79% (n=38), collateral-reported data for 42% (n=20), and clinician-entered data for 17% (n=8). The Modified Colorado Symptom Index was completed by 54% (n=26) of

service users (total score: mean 24.16, SD 16.81). In addition, 35 service users had a support person who could complete the Modified Colorado Symptom Index, and 56% (n=19) of support persons completed it (mean 26.71, SD 14.43).

Conclusions: Implementing UCD principles while developing the Beehive app resulted in early workflow changes and produced an app that was acceptable and feasible for collection of self-reported clinical outcomes data from service users. Additional support is needed to increase collateral-reported and clinician-entered data.

(JMIR Hum Factors 2025;12:e65889) doi:10.2196/65889

KEYWORDS

eHealth; user-centered design; learning health system; psychosis; early psychosis; user-driven development; web-based; data visualization; surveys and questionnaires; measurement-based care

Introduction

Background

Research estimates that the lifetime prevalence of psychotic disorder diagnoses is approximately 1.5%, and the prevalence of psychotic symptoms is between 4.2% and 17.5% [1]. California, the most populous and second most diverse state in the United States [2], had a population of 39.11 million in 2023, according to the California Census Bureau. This indicates that 586,650 to 6.8 million Californians may experience psychosis symptoms in their lifetime. In response to this, many California counties have developed specialty early psychosis intervention (EPI) services, which vary widely in their implementation approach [3]. The Early Psychosis Intervention Network of California (EPI-CAL) [4] was developed to support the provision of quality EPI services and to create an infrastructure to conduct standardized measurement of the impact of early psychosis care delivery. To support this goal, the EPI-CAL team, in collaboration with several California counties, developed a learning health care network (LHCN) consisting of EPI programs across the state. The EPI-CAL LHCN later joined the national Early Psychosis Intervention Network [5], which allowed additional California EPI programs to participate. Members of the LHCN agreed to gather standardized information and outcomes from their clinics as part of measurement-based care (MBC). Collecting this information is critical to support quality early psychosis care provision within clinical programs as well as enhance statewide learning and development. For example, a narrative review of an MBC approach in behavioral health clinics found such benefits as significantly improving clinical outcomes, improving symptoms more quickly, and decreasing treatment costs [6].

To this end, the EPI-CAL team chose to design and implement a web-based and tablet app called Beehive. We chose the name Beehive to reflect the envisioned purpose for the app: to help LHCN programs learn and grow together for the betterment of the collective, just as bees work together to build a hive for the benefit of the colony. Beehive is a robust, stand-alone eHealth app for use by service users, their support persons, and EPI program providers. In this text, we use the term *service user* to refer to the individual with a psychosis diagnosis who is receiving mental health care from an early psychosis program. We use the term *support person* to refer to any person that the service user has chosen to involve in their care. This is typically the individual's parent but might be another family member, a friend, a partner, or some other close relative. Beehive's purpose is to promote MBC in EPI programs by standardizing data collection across a network of programs focusing on community partner priorities; supporting key components of care such as assessment, safety monitoring, and ongoing care delivery; supporting program-level management of care; and aggregating data across a large network to support evaluation and research at state and even national levels [4].

We chose an eHealth approach to implementing MBC due to its appropriateness for the EPI setting and its potential benefits. Despite the perceived challenges related to experiences of suspicion or paranoia, individuals experiencing mental health difficulties, such as schizophrenia and bipolar disorder, have widely found use of eHealth both feasible and acceptable [7-10]. Furthermore, the use of eHealth can support the advancement of MBC. For example, eHealth apps have been previously demonstrated to promote symptom and outcomes monitoring in both early psychosis care [11,12] and LHCNs [13]. Conducting MBC with eHealth enhances its benefits as it allows for data collection to be standardized across programs and instantly available. For example, MBC may promote collaboration across a care team [14-16], which is relevant for EPI programs for which the evidence-based treatment, coordinated specialty care, is inherently team based [17]. Use of eHealth to collect data in this setting allows data collected by one team member or entered by a service user to be instantly available for all team members. eHealth also enhances the benefits of MBC through data aggregation, which enables evaluation of program performance [16,18,19] or promotion of evidence-based treatments [16,20].

Though there are many potential benefits, there are also numerous barriers to implementation of both new eHealth technology and MBC. According to a systematic literature analysis, the top factors posing a barrier to eHealth app implementation include lack of digital health literacy, lack of devices, financing issues, service-user cognition, and security [21]. These barriers contribute to low adoption and user acceptance, which limit the success of implementation [22,23]. Barriers to implementation of MBC include training burden, concern that negative feedback causes harm to service users, and the time required for survey completion [24-26].

To pursue the benefits of using eHealth to implement MBC and mitigate the potential barriers, we developed Beehive with user-centered design (UCD) principles. UCD prioritizes the needs and expectations of the end user [27,28]. UCD approaches include dedicated design activities, active involvement of users

in the design process, incorporation of their feedback, prototyping, and continuous evaluation [29], which can address low user acceptance and low adoption [30]. Beehive's iterative development began with a collaborative process with service users, support persons, and EPI program providers to identify and prioritize which outcome measures should be collected in the app [31]. We also explored how service users, support persons, and EPI program providers wanted to be informed about data-sharing options in Beehive and built Beehive's end user license agreement (EULA) workflow to incorporate user perspectives [32]. With the survey content and EULA workflow finalized, we moved onto the development of the user-facing parts of Beehive.

Objectives

In this study, we aimed to (1) use UCD principles to create a co-designed web-based and tablet app, called Beehive, to support MBC in EPI programs, and (2) assess Beehive's initial feasibility in clinical settings by piloting it in 4 EPI programs.

Methods

Design

To promote Beehive engagement across multiple types of users and across multiple domains of engagement, we integrated UCD principles of incorporation of user feedback, prototyping, iterative design, and continuous evaluation. Service users, support persons, and EPI program providers had multiple opportunities to provide feedback, which was incorporated throughout Beehive development. Figure 1 shows the study design from conceptualization through data collection.

Figure 1. Study design for Beehive development.

The development process began with conceptualization of user journeys. User journey mapping envisions how specific types of users, such as a service user or a service provider, will interact with an app from access point through all required activities [33]. User journey mapping allowed us to identify which storyboards we should develop to present in UCD workshops and which user-types we needed to recruit for those workshops. Storyboards are a tool to visualize app workflows and the user interface [34]. We developed storyboards to present as prototypes in UCD workshops so that feedback could modify the app design before time was invested in coding the alpha version of the app. The alpha version of the app included core workflows and was both evaluated internally and presented in another UCD workshop to gather more feedback before coding the beta version of the app. The beta version of the app incorporated remaining feedback from storyboard workshops, new feedback from the alpha stage, and added the remaining core functionality that was not in-scope for the alpha version (eg, reports). The beta version of Beehive was piloted by 3 EPI programs over 6 months to further refine the app before launching it across all EPI-CAL programs. We used pilot data to assess initial use and uptake of Beehive's beta version.

This UCD approach allowed the EPI-CAL team to receive and incorporate feedback during conception, design, and testing phases of eHealth app development, and include multiple perspectives to facilitate user engagement in eHealth. Notably, UCD has also been demonstrated to increase eHealth adoption and user acceptance in research and clinical settings [30,35].

Participant Recruitment

UCD Workshops

For UCD workshops of the Beehive storyboard, we recruited participants from the following three EPI community partner

groups: (1) EPI service users, (2) their support persons, and (3) EPI providers. Eligible participants were (1) actively or formerly affiliated with an EPI-CAL EPI program, (2) English speaking, and (3) able to provide written informed consent and assent (minors or conserved adults). EPI program providers were recruited through contact with the team lead of the 12 active EPI-CAL EPI programs. Service users and support person participants were invited either through EPI program provider referral or by the research team directly contacting individuals who had previously consented to be contacted for future research opportunities.

One EPI-CAL clinic agreed to participate in a workshop for the alpha version of the app to support the refinement of Beehive before piloting.

Piloting

In total, 4 EPI-CAL clinics agreed to participate in 6 months of Beehive beta testing before Beehive's full launch across the entire EPI-CAL LHCN. During this period, programs integrated the Beehive app into standard clinical care. Pilot sites registered service users who were active in their program at the time of launching Beehive and new service users who started after the launch. Each participating program has different acceptance criteria for service users, and this has been described in a separate protocol paper for the EPI-CAL study (NCT04007510) [4]. Service users and support persons were excluded from piloting if they did not speak English because Beehive beta version was only available in English. At their first point of contact with Beehive, service users and primary support persons completed the Beehive EULA and were asked if they gave permission for their clinical data collected in Beehive to be used for research purposes [32]. We trained clinics to involve the legal guardians of service users aged <18 during EULA completion, and these service users were required to have a primary support person registered in Beehive. Individuals could update their data-sharing permissions at any time.

Methods

User Journey Mapping

The EPI-CAL research team worked collaboratively with the contracted app developer in the user journey mapping and storyboard design phases for the Beehive tablet and web apps. Three primary user groups with distinct roles were identified: (1) service users or support persons, (2) direct-service providers, and (3) program administrators. Beehive user journeys were developed for each group.

For all user groups, user journeys were designed to guide individuals smoothly through Beehive onboarding and account creation to the EULA explainer video detailing the types of information Beehive collects, who can access their data, and how to select their preferred permissions for who can view their data [32]. Users then choose their data-sharing permissions.

Beehive then presents service users and support persons with a series of one-time and longitudinal surveys to measure clinical outcomes at 6-month intervals. Specific survey items associated with risk, such as suicidal or homicidal ideation, send real-time alerts to EPI program providers if they are endorsed by service users. Beehive creates visualizations of this survey data. While service users and support persons cannot independently access survey visualizations, we considered this is as part of their user journey because it should be shown to them by EPI program staff as part of regular care.

EPI program provider user journeys were designed to facilitate easy management of service user records and smooth navigation to service user data. A dashboard presents the most important information to users, such as outstanding survey alerts or other action items. A client list presents all registered service user records in list format with the most relevant information displayed. EPI program provider users can click into service user records to view survey results, survey visualizations, and complete provider-entered surveys. Providers can display survey results and survey visualizations as part of ongoing care with service users and their support persons to facilitate understanding and coordinate treatment priorities.

Administrator user journeys were designed to promote clinic-level management tasks, such as Beehive implementation and quality assurance. Administrator dashboards present aggregate-level information of survey data and allow for the ability to compare clinic averages to the average of the EPI-CAL LHCN. Administrator dashboards also present summaries of Beehive activity across the clinic. Administrator users can also download reports, including survey data reports, to use their clinic data for quality assurance or reporting requirements.

UCD Workshops

Next, the development team created dynamic storyboards of the above user journeys. These storyboards were presented to community partners from early psychosis clinics in 90-minute UCD workshops. Figures 2 and 3 are images from the storyboards presented to community partners.

In the storyboard workshops, we presented major features of the app and asked for feedback on the app's look and feel, as well as functionality as it related to existing clinical workflow, and ease of use and acceptability for service users, their support persons, and EPI program providers. Because we determined 3 user-types during user journey mapping, we held 3 different types of storyboard workshops tailored to the journeys of these users: service users or support persons, direct-service providers, and program administrators. If EPI providers had roles in both direct-service provision and program administration, they could attend both groups. UCD workshops were conducted through videoconferencing to comply with the COVID-19 social distancing restrictions. Each session was audio recorded and included 2 facilitators (KEB, LMT, TAN, or SE) and a notetaker (VLT). There were no other individuals present other than researchers and participants. The notetaker took detailed notes, as close to verbatim as possible. Audio recordings were used as a reference to fix any portions of the notes which were not clear.

After storyboard workshops, we integrated feedback to make design changes to the alpha version of the app. During a 90-minute alpha testing workshop, we solicited feedback on the alpha app, with a special emphasis on how compatible it was with the existing clinical workflows. We created test

accounts for each participant and had them complete various core workflows in the app, such as registering a service user, completing surveys, and reviewing data visualizations. This workshop included one facilitator (KEB) and one notetaker (LS). After the conclusion of all workshops, we continued to integrate feedback to make design changes in the beta version of the app.

Figure 2. Survey item in storyboard.

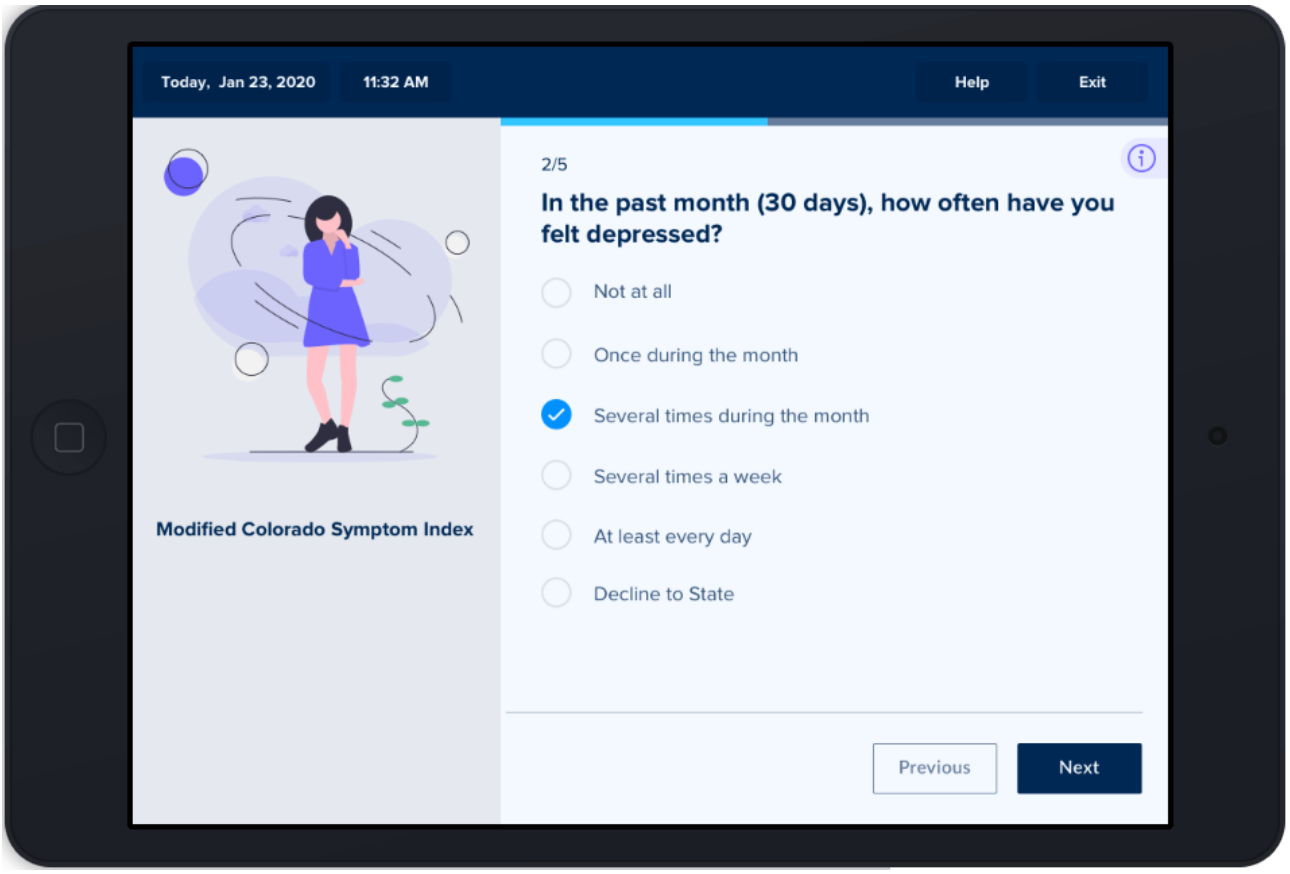
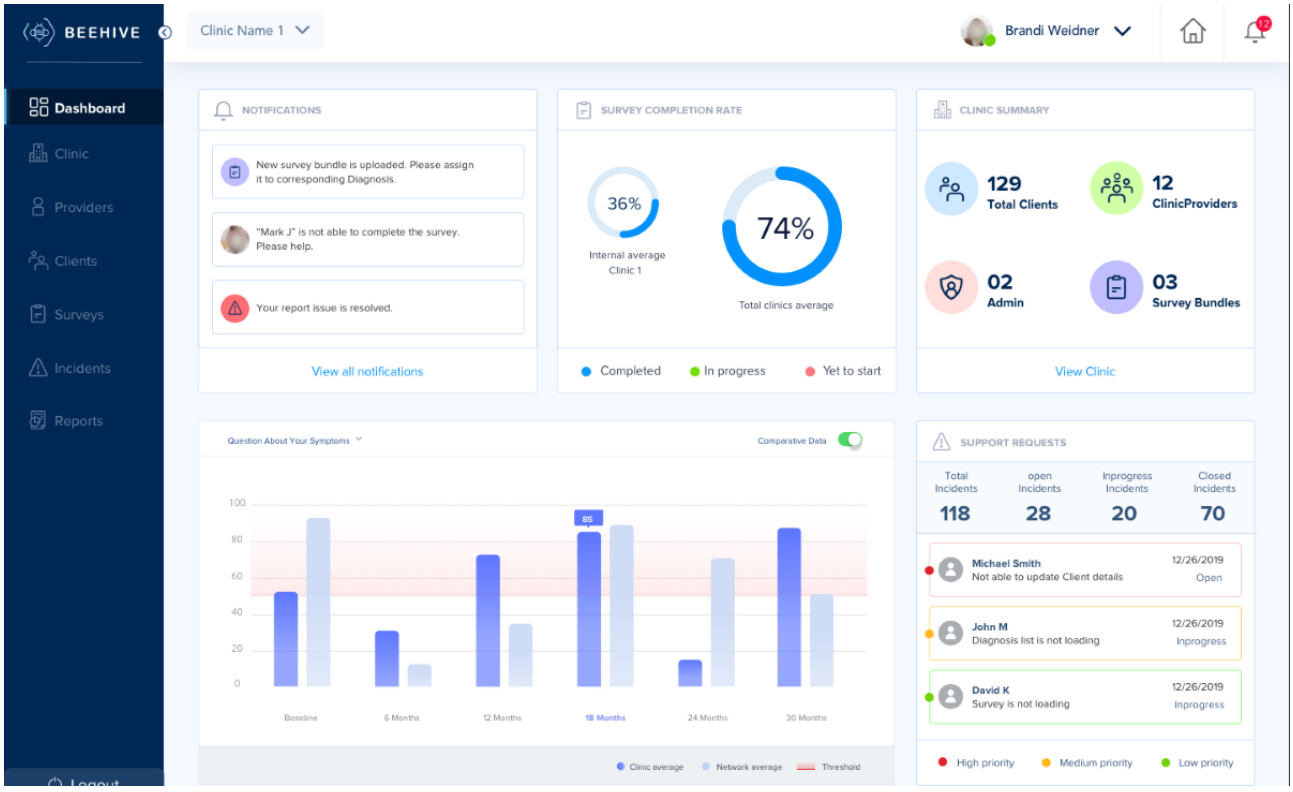


Figure 3. Clinical administrator dashboard in storyboard.

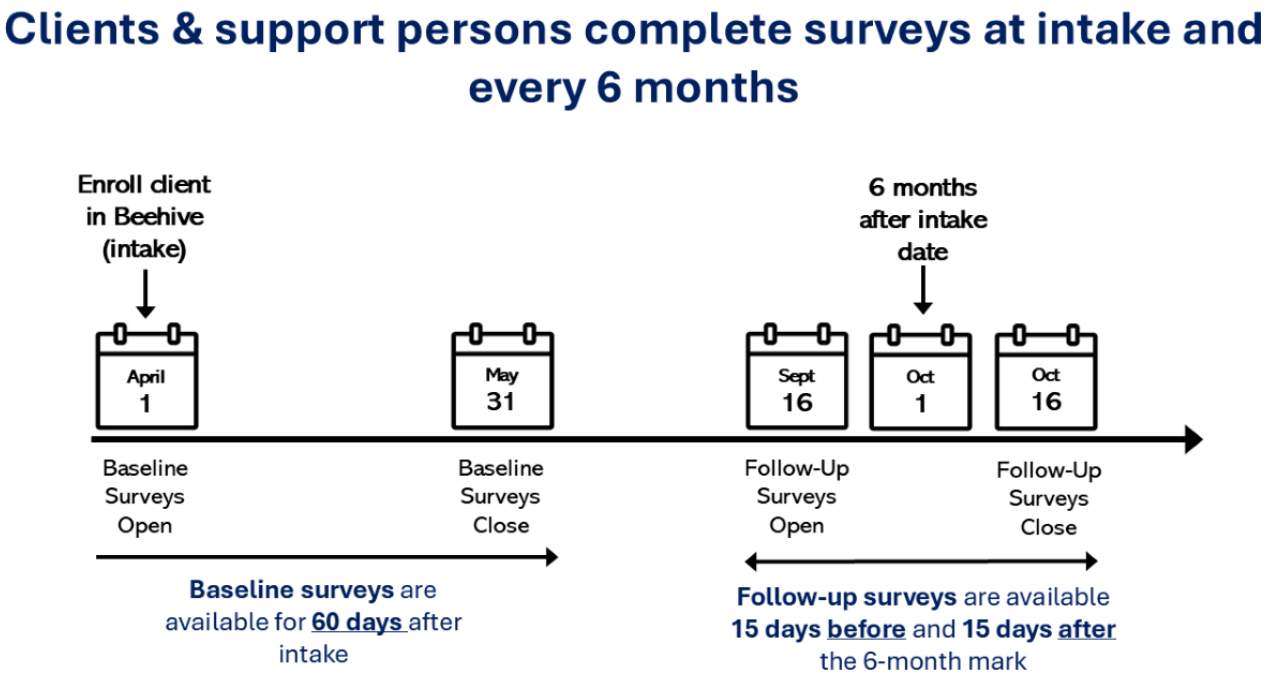


Piloting

Finally, piloting of the beta version of Beehive was conducted over a 6-month period with EPI-CAL LHCN programs that were identified as pilot sites. We provided training to each pilot site and assigned them a point person from our research team to provide regular support, troubleshoot implementation challenges, and escalate app bugs and implementation barriers. We trained all program staff, regardless of clinical role. The training series showed users how to complete key Beehive workflows, introduced the EPI-CAL core assessment battery, and included activities on how to interpret data visualizations. We also met with key staff at each program to support them to devise a plan to integrate Beehive into their existing clinical workflows, such as how to integrate service user registration during the clinical intake process. The full description of this training and support is described in a separate paper [4]. During this training and support process, we received informal feedback from program staff about implementation successes and challenges.

During piloting, programs integrated the Beehive app into standard clinical care for all service users. Service users and their support persons were registered in Beehive web app by staff at their EPI program. Program staff entered the service users' treatment start date which was considered the start of the service users' baseline window in Beehive. At enrollment, baseline, and every 6 months thereafter, surveys were assigned to service users, their support person, and their provider. One-time surveys assessing lifetime experiences were available for all respondents at enrollment. Service users had 3 enrollment surveys, support persons had 1, and clinicians had 1. Longitudinal surveys were assessed at treatment baseline and every 6 months thereafter. Service users had 17 longitudinal surveys, support persons had 6 surveys, and clinicians had 9 surveys. The duration of baseline survey window was 60 days; the duration of follow-up survey windows was 30 days (15 days before and after the target completion date). Figure 4 shows a Beehive training slide with a visualization of survey windows.

Figure 4. Beehive training slide showing survey windows during piloting.



Programs were instructed to enroll all service users, regardless of how long they had been affiliated with the EPI program. Therefore, some individuals may have been enrolled in Beehive after their baseline window had closed, and their first time point during piloting may have been a 6-month time point or a 12-month time point. Service users and support persons could access and complete surveys in-person at the clinic on the tablet app or they could complete them remotely via "web link." The "web link" was a unique link that was texted or emailed to them weekly during survey windows if surveys were not fully completed. Surveys accessed via web link could be completed on any personal device that had access to the internet and a web browser. Service providers completed surveys and could review data on the web app. Baseline surveys were intended to support the clinical intake process, including initial assessment and collaborative treatment planning. Follow-up surveys were intended to support ongoing assessment, adjustments to treatment planning, and monitoring of treatment goals.

The EPI-CAL core assessment battery, including how it was created and all included measures, is described in a separate paper [4]. Briefly, Beehive survey content includes both the Early Psychosis Intervention Network core assessment battery [36] and additional measures based on EPI community partner feedback determined in earlier qualitative work for the EPI-CAL study [4,31]. A table of measures is included in the Multimedia Appendix 1 [4]. One survey of particular interest is the Modified Colorado Symptom Index (MCSI) [37], a measure central to the aims of the broader EPI-CAL study and which we asked sites to prioritize [4]. The MCSI is a 14-item, self-report scale

which measures the frequency of psychiatric symptoms, including symptoms of mood, psychosis, cognition, forgetfulness, and risk to self and others. Respondents indicate frequency of symptoms over the past 30 days on a 0 to 4 scale of "not at all" to "at least every day." Total scores range between 0 and 56, with higher scores indicating higher frequency and number of psychiatric symptoms.

Data Analysis

In UCD workshops, we asked highly structured questions to solicit feedback on the storyboard and alpha version of Beehive. Subsequently, we organized our data categories relevant to the workflows and features we were evaluating. We then organized comments by whether they were supportive of the existing features or critical and asked for change so that we could focus on what features to move forward and what features to change as we created the alpha and beta versions of Beehive.

To investigate the initial feasibility of Beehive in EPI-CAL clinics, we reviewed descriptive statistics of pilot participants, including registration, enrollment, participant characteristics, and survey completion. Engagement with surveys and survey completion were examined in three ways as follows: (1) determine the proportion of service users for whom any data were entered, regardless of respondent type, (2) determine the rate of survey completion across all available surveys, and (3) evaluate whether participants completed all, partial, or no surveys across survey time points during piloting phase. Partial survey completion indicates that the respondent completed at least 1 survey, but did not complete all of their surveys in the specified time point. We also evaluated completion of MCSI because it is a measure that we asked programs to prioritize.

Ethical Considerations

The institutional review board of the University of California, Davis, approved the study (1403828-21, California Collaborative Network to Promote Data-Driven Care and Improve Outcomes in Early Psychosis). In addition, several of the counties and universities with a program participating in EPI-CAL required a separate review and approval of the project by their institutional review board. All study participants provided written informed consent and assent (as appropriate). Participants received US $30 compensation for each workshop they participated in. Participants in the piloting phase were not compensated because integration of Beehive was part of routine care in the EPI program. Audio recordings UCD workshops include voice print identifiers and are stored in compliance with University of California, Davis HIPAA (Health Insurance Portability and Accountability Act) policies and procedures. Data collected during piloting for research includes limited identifiers including zip code, dates of service, and month and year of birth. Only trained research staff with a need-to-access have access to identifiable data.

Results

UCD Workshops

We conducted 14 storyboard workshops with 77 total participants between April 3, 2020, and August 28, 2020. In total, 4 workshops were with service users (n=8, 10%) and their support persons (n=9, 12%). In addition, 10 workshops were with EPI program providers (n=60, 78%), including 6 for service providers, 3 for administrators, and 1 for both service providers and administrators. Demographics for workshop participants are provided in Table 1.

We completed an interim analysis of storyboard workshop data in May 2020. We completed the final analysis of workshop data in August 2020, after all groups were completed. After each analysis, we discussed and synthesized the feedback for the developers to support app development. We attempted to balance the needs of all types of participants. However, if there were needs or feedback in direct contrast with one another, we prioritized service-user feedback due to our value of centering service-user feedback in this app. This feedback and the action taken to address it are summarized in Table 2.

We conducted 1 alpha workshop in October 2020 with 4 EPI program provider participants. Feedback from this workshop was analyzed in October 2020. During this workshop, participants identified a few bugs in the app, but their feedback primarily focused on ideas for integrating Beehive into clinical workflows. For example, they suggested that Beehive training should include best practices for how providers can review the data, engage with the data, and make the most out of Beehive. They also shared concerns about using technology in telehealth settings. For example, switching to telehealth in response to the COVID-19 pandemic had been very difficult for some families, and they predicted those same families would find using Beehive challenging if the clinic could not meet with them in-person to teach them how to use it. They were less concerned about service users and support people using Beehive on a tablet in the clinic where they could provide in-person support. Finally, participants brought up the importance of shifting the culture of clinics to view data collection as an important part of treatment, not just an extra task where information is being extracted from service users. For example, surveys should be directly related to service-user recovery goals. Participants discussed how visualizations could be used to demonstrate the clinical utility of gathering these data. For example, 1 clinician said they would want to use the graphs to point out the way a service user is improving or doing better and that they would want to highlight their strengths. Another participant cautioned that some service users may not want to look at data visualizations and that this should be an optional part of their care. Figures 5 and 6 illustrate design changes present in the beta version of app after the conclusion of all workshops.

Table 1. Demographics of user-centered design workshops[a].

	Service users (n=8)	Support persons (n=9)	EPI[b] program providers (n=60)
Clinic type, n (%)			
Medi-Cal	7 (88)	6 (67)	30 (50)
Private insurance	<5 (<63)	<5 (<55)	30 (50)
Age (y), mean (range)	22.50 (16-33)	41.50 (14-60)	36.25 (26-50)
Sex at birth, n (%)			
Female	<5 (<63)	8 (89)	43 (72)
Male	6 (75)	<5 (<55)	17 (28)
Gender, n (%)			
Female	<5 (<63)	8 (89)	43 (72)
Male	6 (75)	<5 (<55)	16 (27)
Nonbinary	<5 (<63)	—[c]	—
Missing	—	—	<5 (<8)
Race, n (%)			
African or African American or Black	<5 (<63)	—	5 (8)
Asian	—	—	6 (10)
White or Caucasian	<5 (<63)	6 (67)	33 (55)
Other	<5 (<63)	<5 (<55)	10 (17)
More than one	<5 (<63)	<5 (<55)	<5 (<8)
Missing	—	<5 (<55)	<5 (<8)
Ethnicity, n (%)			
Latinx	<5 (<63)	5 (56)	26 (43)
Not Latinx	5 (63)	<5 (<55)	33 (55)
Missing	—	—	<5 (<8)
Sexual orientation, n (%)			
Bisexual or gay or lesbian	<5 (<63)	—	5 (8)
Heterosexual	5 (63)	9 (100)	54 (90)
Other	<5 (<63)	—	<5 (<8)

[a]Cells with fewer than 5 individuals are masked to protect the identity of participants.

[b]EPI: early psychosis intervention.

[c]Not available.

Table 2. Implementation of feedback from user-centered design workshops.

Problem or need identified in workshop	Solution implemented in alpha version
The color scheme and layout seemed "overly clinical"	Brought in more color into the palette and added icons for visual information
Some important aspects of the user-interface were too subtle, such as the survey progress bar or the urgent clinical issues widget	Changed color and design to make it more prominent
Service-user and support person registration were only available as self-registration and could not be completed by EPI program providers	Added this workflow to the web app so that EPI program providers may complete it in advance of service users engaging with surveys
Clinic-level data for service-user demographics was not visualized	Added clinic-level visualizations for race, ethnicity, sex, gender identity, and other demographic metrics
Not all service users or support persons wanted to see score thresholds or comparative data on clinical measures, but some did	Added a toggle to individual-level visualizations so that users can turn on the threshold information or comparative data if they want to see it, or turn it off if they do not want to see it
Service users might have differed on which individual-level survey visualization they wanted to see by default on their data view page	Added feature that allows users to set which measure displays by default for each service user
Some language used in the app needed to be clarified for users to understand what data was being collected or how certain features worked	Changed "homeless" to "without a permanent address" when assessing housing status; changed "help" to "Ask for help" to make it clearer that selecting button will alert the EPI program provider; changed "Diagnosis" to "Primary diagnosis"
Different programs used different words to refer to service users, and individuals might have varied on their preference for what word to use regardless of what their program tended to use	Wherever possible, implemented dynamic text so the service user's preferred name shows throughout the app, rather than any specific word to denote "service user"
Program staff wanted to see overall progress on completion of all surveys at any given time point	Added in visual indicator to show survey completion across multiple surveys (not just while completing one individual survey)
When visualizing a survey, they wanted to have more than just the global score visualized. Also, they wanted a visualization that showed responses to individual items	Added in a visualization that shows individual items as well as the global score
Service users and support persons might not have preferred the official names for measures and might have preferred a more simplified title	Added the ability to enter a display name for surveys (eg, "Family Impact" instead of "Burden Assessment Scale")
Early psychosis intervention program providers needed a way to see both the official measure name as well as the display name	Added a hover modal on survey titles to show the display name for the survey
The provision of clinic services might have been fully remote for the foreseeable future, and the current design of Beehive only allowed service users and support persons to complete surveys on a tablet in-person at the clinic	Design a web link solution which allows service users and support persons to answer surveys remotely. A link to complete their surveys can be emailed or texted to them

Figure 5. Survey item in beta.

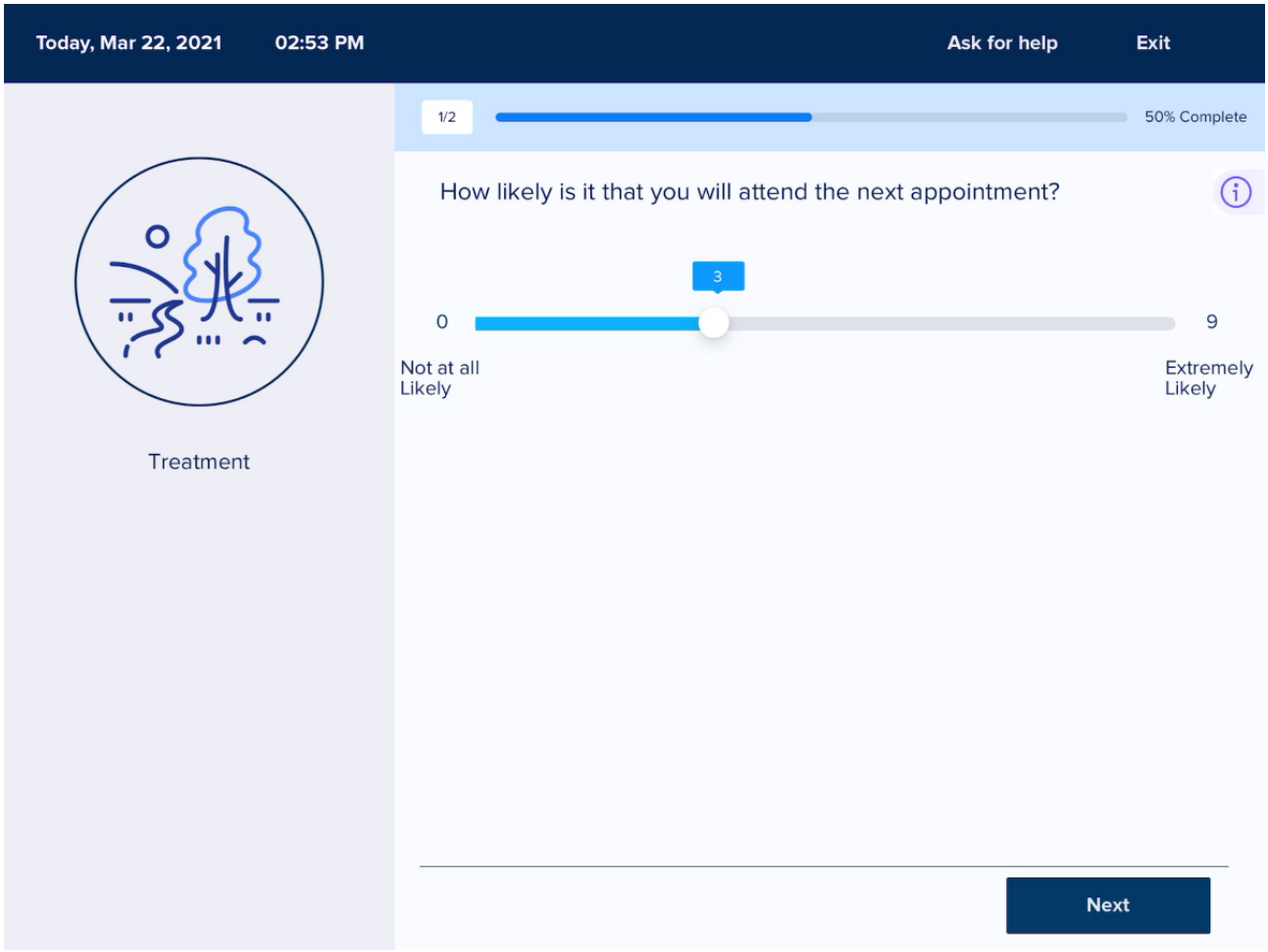
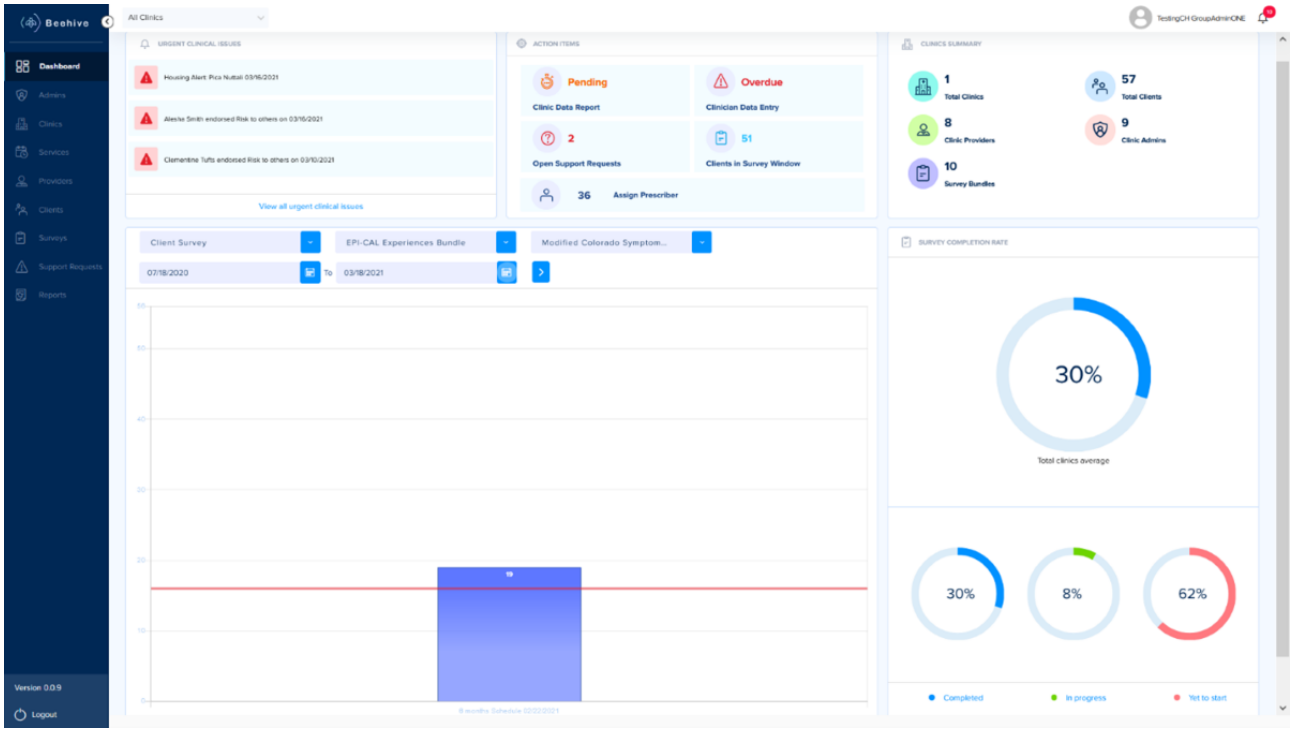


Figure 6. Clinical administrator dashboard in beta.



Piloting

We conducted piloting of Beehive beta app between March 2021 and September 2021. Our training and ongoing support of pilot sites allowed us to gather informal feedback about both the training and the Beehive app. We used this feedback to make adjustments in real time, when possible, or to plan for future changes to Beehive.

We made real-time changes to the training approach in response to program needs based on our observations and their feedback. At the time of training pilot sites in early 2021, these EPI programs were navigating constant uncertainty related to the COVID-19 pandemic, including influx of service users, uncertainty about work location, reduced workforce, etc. In response to this environment, we found it necessary to ask sites to focus on small implementation steps even though we trained them on all available workflows. For example, we asked pilot sites to initially focus on engaging service users and their support persons to complete enrollment and complete surveys. When that was mastered, we asked them to focus on engaging new service users and support persons with Beehive during their clinical intake process. Finally, toward the end of the piloting period, we encouraged them to focus on entering EPI program provider-entered data. Even if pilot sites registered existing service users, they were asked to set the service user's survey baseline date to align with their start in the EPI program. Therefore, some service users may have never been assigned surveys during their baseline survey window.

During the piloting phase, 93 service users, 78 support persons, and 86 service providers were registered across 4 clinics. Of the 93 service users who were registered into Beehive by their program, 59 (63%) completed the Beehive EULA during piloting, including 48 (51%) individuals who gave permission to use their data for research. Of the 78 support persons registered to a service user in Beehive, 52 (66%) completed the Beehive EULA, including 42 (54%) individuals who gave permission to use their data for research. Of these 42 support persons, 5 were excluded from analysis because the service user they were registered with did not give permission to use data in research, and we prioritize the data-sharing decision of the service user regarding use of collateral data for research purposes.

While most users entered at least one survey window during the piloting phase, 3 were discharged from their program before longitudinal surveys were available. Of the 86 service providers registered, 78 (91%) completed the Beehive EULA, including 72 (84%) individuals who gave permission to use their data for research. This information is presented in Figure 7.

Participant demographics and EPI program provider professional background are provided for individuals who agreed to share their data for research purposes in Tables 3 and 4. Of note, age is missing for some EPI program providers and support persons. This data was collected during registration, and this field was not included in the first release of the beta version of the app. Therefore, some users were not able to complete this field during registration and did not return later to update it, yielding missing data for 15 (36%) support persons and 37 (51%) EPI program providers.

First, to examine engagement during piloting, we assessed the number of service users that had entered any data that could be used in care. During piloting, respondents entered survey data for 85% (n=41) of service users. This includes self-reported data for 75% (n=36), collateral-reported data for 42% (n=20), and EPI program provider–entered data for 17% (n=8).

Second, we evaluated survey completion across the total amount of available surveys. A total of 1517 surveys were assigned across all respondent types during piloting and 35.4% (n=537) of those surveys were completed across all time points. We also evaluated survey completion by respondent type. Across all time points, service users completed 396 (49.4%) of 802 assigned surveys. Support persons completed 113 (47.5%) of 238 assigned surveys. EPI program providers completed 28 (5.9%) of 477 assigned surveys.

Finally, we evaluated how many respondents completed all, partial, or no surveys at each time point. Because participants could be enrolled at any point in treatment, the first time point for a service user may not have been their "baseline" appointment. These data are presented in Table 5.

The MCSI was completed by 54% (26) of service users, and 7 were excluded for responses of "prefer not to say" (Total Score: mean 19.58, SD 16.81). Additionally, 35 service users had a support person who could complete the Modified Colorado Symptom Index, and 56% (n=19) of support persons completed it. From these 19, 5 were excluded for responses of "prefer not to say" (mean 26.71, SD 14.43).

Through piloting, we also gathered informal feedback about workflows that could be improved in Beehive and that we would address with future change orders to the app after the testing of the beta version of Beehive. For example, we received feedback that the survey windows, initially chosen to mirror the data collection windows of clinical trials, were far too narrow and restrictive for data collection in community mental health programs. Program staff users also indicated that they wanted a better way of seeing a summary of what surveys service users and support persons completed. Service user and support person feedback was relayed to our team via program staff. For example, service users and support persons wanted to customize the day and time they received the web link via SMS text messaging and email.

Figure 7. Registration and enrollment during Beehive piloting.

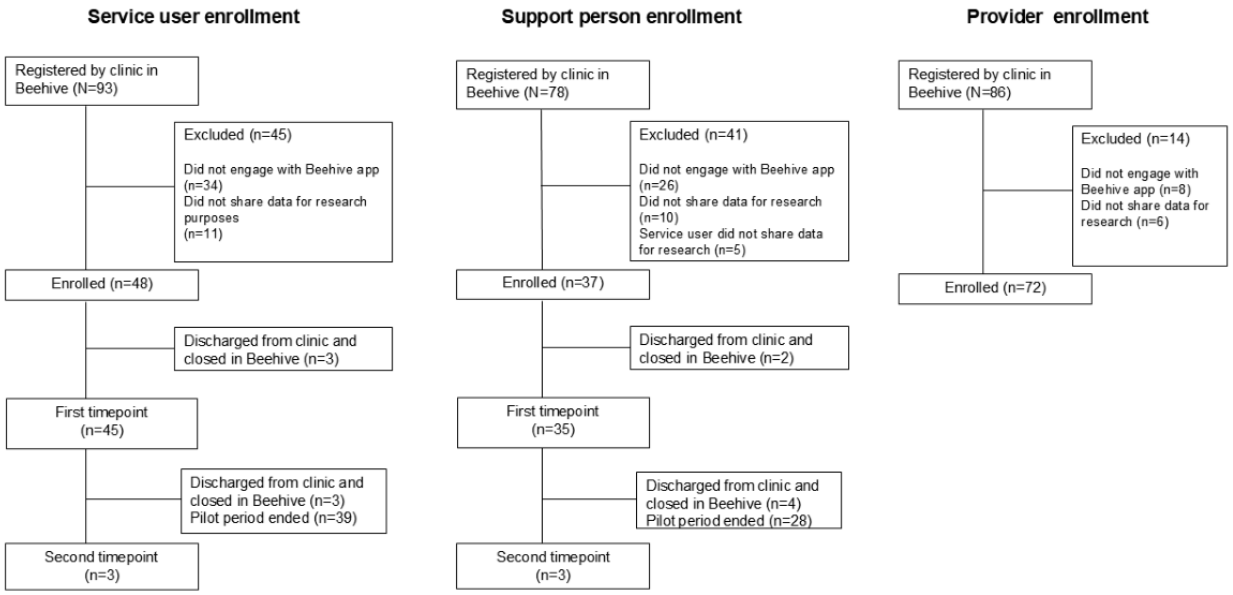


Table 3. Demographics of Beehive pilot participants[a].

	Service users (n=48)	Support persons (n=42)	EPI[b] program providers (n=72)
Age (y), mean (range)	18.88 (12-31)	44.11 (31-61)[c]	33.89 (22-50)[d]
Sex at birth, n (%)			
Female	24 (50)	21 (70)	30 (42)
Male	24 (50)	6 (20)	5 (7)
Prefer not to say	__[e]	<5 (<12)	—
Missing	—	—	37 (51)
Gender, n (%)			
Female	18 (38)	21 (70)	30 (42)
Male	22 (46)	6 (20)	5 (7)
Nonbinary	<5 (<10)	—	—
Questioning or unsure of gender identity	<5 (<10)	—	—
Prefer not to say	<5 (<10)	<5 (<12)	37 (51)
Race, n (%)			
African or African American or Black	15 (31)	7 (17)	5 (7)
American Indian or Alaskan native	<5 (<10)	—	—
Asian	<5 (<10)	<5 (<12)	8 (11)
Hispanic or Latinx only	10 (21)	<5 (<12)	19 (26)
White or Caucasian	10 (21)	10 (24)	33 (46)
More than one race	8 (17)	<5 (<12)	5 (7)
Other	—	—	<5 (<7)
Prefer not to say	<5 (<10)	<5 (<12)	—
Unsure	—	<5 (<12)	—
Missing	—	12 (29)	<5 (<7)
Ethnicity, n (%)			
No—I do not identify as Hispanic or Latinx	27 (56)	17 (40)	41 (57)
Yes—I identify as Hispanic or Latinx	14 (29)	5 (12)	28 (39)
Unsure or do not know	5 (10)	5 (12)	—
Prefer not to say	<5 (<10)	<5 (<12)	<5 (<7)
Missing	—	12 (29)	<5 (<7)
Service user diagnosis, n (%)			
Clinical high risk	—	—	—
Attenuated psychosis symptoms	6 (13)	—	—
First episode psychosis	—	—	—
Substance induced psychotic disorder with onset during intoxication	<5 (<10)	—	—
Mood disorders with psychotic features	9 (19)	—	—
Schizoaffective disorder (bipolar or depressive type combined)	10 (21)	—	—
Schizophrenia	5 (10)	—	—
Other specified schizophrenia spectrum disorder	<5 (<10)	—	—
Unspecified psychosis	<5 (<10)	—	—
Other first episode psychosis	7 (15)	—	—

	Service users (n=48)	Support persons (n=42)	EPI[b] program providers (n=72)
Clinical high risk or first episode psychosis status not confirmed	—	—	—
Anxiety disorders	<5 (<10)	—	—
Number of support persons registered in Beehive, n (%)			
None	14 (29)	—	—
1	29 (60)	—	—
2	5 (10)	—	—
Relationship of support persons with service user, n (%)			
Parent (adoptive)	—	5 (12)	—
Parent (biological)	—	34 (81)	—
Stepparent	—	<5 (<12)	—
Spouse or partner	—	<5 (<12)	—
Sibling	—	<5 (<12)	—

[a]Cells with less than 5 individuals are masked to protect the identity of participants.

[b]EPI: early psychosis intervention.

[c]Data missing for 15 individuals.

[d]Data missing for 37 individuals.

[e]Not available.

Table 4. Professional background of early psychosis intervention program providers registered during Beehive pilot (N=72)[a].

Background	Values, n (%)
Education level	
HS[b] diploma or GED[c]	7 (10)
Associate's degree	<5 (<7)
BA[d] or BS[e]	16 (22)
MA[f] or MS[g]	18 (25)
MFT[h]	7 (10)
MSW[i]	<5 (<7)
PsyD[j]	5 (7)
PhD[k]	6 (8)
MD[l]	9 (13)
Professional role	
Administrative support staff	<5 (<7)
Case manager or recovery coach	<5 (<7)
Clinic coordinator	6 (8)
Clinical supervisor or team lead	<5 (<7)
Clinician or therapist	30 (42)
Family advocate	<5 (<7)
Peer support specialist	<5 (<7)
Prescriber or psychiatrist or Other medical personnel	9 (13)
Program director	<5 (<7)
Research staff	<5 (<7)
Supported education and employment specialist	<5 (<7)
Other	7 (10)
Licensure status	
Unlicensed	49 (68)
Licensed	23 (32)
Years licensed (n=23)	
≤1	8 (38)
1 to 6	7 (33)
≥7	6 (29)
Number of languages in which services are provided	
1	47 (65)
2	18 (25)
Missing	7 (10)
Languages for service provision[m]	
English	60 (92)
Spanish	18 (25)
Arabic	<5 (<7)
Hmong	<5 (<7)
Tagalog	<5 (<7)

Background	Values, n (%)
Other	<5 (<7)

[a]Cells with less than 5 individuals are masked to protect identity of participants.
[b]HS: high school.
[c]GED: general educational development.
[d]BA: Bachelor of Arts.
[e]BS: Bachelor of Science.
[f]MA: Master of Arts.
[g]MS: Master of Science.
[h]MFT: Master of Marriage and Family Therapy.
[i]MSW: Master of Social Work.
[j]PsyD: Doctor of Psychology.
[k]PhD: Doctor of Philosophy.
[l]MD: Doctor of Medicine.
[m]Respondents could select more than one response, so percentages will be greater than 100%.

Table 5. Survey completion by respondent type[a].

	Service users, n (%)	Support persons, n (%)	EPI[b] program providers, n (%)
Survey completion at enrollment[c]			
All	30 (63)	17 (50)	10 (21)
Partial	5 (10)	__[d]	—
None	13 (27)	17 (50)	38 (79)
Survey completion at first time point[e]			
All	18 (40)	17 (55)	—
Partial	8 (18)	3 (10)	4 (9)
None	19 (42)	13 (42)	41 (91)
Survey completion at second time point[f]			
All	1 (33)	—	—
Partial	—	—	—
None	2 (67)	2 (100)	3 (100)

[a]Partial survey completion indicates that respondents completed at least one survey, but did not complete all assigned surveys.
[b]EPI: early psychosis intervention.
[c]Total respondents for service users, n=48; support persons, n=34; and service providers, n=48.
[d]Not available.
[e]Total respondents for service users, n=45; support persons, n=31; and service providers, n=45.
[f]Total respondents for service users, n=3; support persons, n=2; and service providers, n=3.

Discussion

Principal Findings

This study describes the EPI-CAL program's design and acceptability testing approach for a custom web-based and tablet app, Beehive, to support systematic data collection, care delivery, program evaluation, and research across a statewide network of EPI programs. Our goal was to develop an app that was clinically useful for, usable by, and acceptable to diverse EPI programs across the state of California.

To ensure the app best matched the needs of the EPI participants, we adopted a UCD approach to develop Beehive. Previous research in the mental health digital space supports that active involvement from the app's intended users during the development phase can improve the appropriateness of the end product for the users of interest [38]. Initial feedback across the 3 development phases was primarily collected in workshops (storyboard and alpha version) and during pilot implementation (beta version). In storyboard and alpha workshops, we presented prototypes to demonstrate major features of the app and asked for feedback on the app's "look and feel," compatibility with existing clinical workflow, and ease of use and acceptability for service users, their support persons, and EPI program providers. Consistent with other studies who have included end users during the design phase of their eHealth apps [38-41],

feedback from these workshops resulted in immediate changes to the alpha and beta apps that would not have otherwise been made.

During piloting, we continued to collect user feedback around Beehive features, as well as assess acceptability of the app by examining preliminary enrollment and survey completion. Our enrollment and survey completion rates are consistent with the acceptability of other mental health apps developed using a UCD approach [42,43], although there is wide variability depending on the implementation approach.

During the design and testing phase, we observed that different types of community partners expressed different, and at times conflicting, needs. For example, we asked participants about their preferences for seeing score thresholds or comparative data as part of the visualization for their clinical measures. Some service users said that, in times of relapse or increasing intensity of symptoms, additional information on the visualization would be demoralizing. In contrast, many participants could imagine scenarios where that information would be useful as a form of psychoeducation to normalize service-user experiences or understand the relative severity of symptoms. To address these diverse needs and promote engagement with Beehive we added a toggle to individual-level visualizations so that users can turn the threshold information or comparative data on or off. A flexible design approach that is tailored to an individual's needs has been shown to be more efficacious in a mobile health app setting [43,44]. Therefore, design changes incorporated flexibility where possible to enable our team to meet the various needs of individuals while maintaining consistent implementation to meet evaluation and research goals.

Similarly, user feedback informed our training approach. For example, some EPI program provider users expressed that they would use the graphs in the app in clinical care with service users to highlight strengths and progress. In contrast, another EPI program provider participant cautioned that some service users may not want to look at data visualizations and that this should be an optional part of their care. Thus, our training approach highlights how visualizations may be used in direct care but are not prescriptive. Feedback from workshop participants also highlighted the importance of shifting clinic culture to view data collection as a key part of care provision. Our team considers EPI program providers to be integral in promoting engagement for service users and their support persons as it is the EPI program providers who communicate why Beehive is being used in care. To begin addressing potential barriers of buy-in and engagement for all users, we designed our trainings to include the context of why Beehive and MBC were being implemented in their programs, including a presentation on the potential value of Beehive (designed and delivered by author LS) [45].

During piloting, we observed barriers to integration. For example, despite our designing the first training such that the programs could start registering service users immediately afterward, the programs failed to do so. When asked, programs informed us that they were nervous to receive questions about Beehive that they did not know how to answer. This resulted in our team creating materials to provide more structure for programs as they introduced Beehive to service users, such as an introduction script, Beehive infographics, and other handouts. Once programs started enrolling existing service users, many found it difficult to transition enrolling new service users, given that they already had numerous documentation requirements during their clinical intake process. In response, we added a "workflow meeting" to our training series where we asked program leadership and key staff involved in intakes to walk us through their existing procedures so that we could help brainstorm where the required Beehive workflow steps could be implemented and who from their program would be responsible for each step. We additionally observed that clinician-entered data were hard for sites to prioritize. For example, there was a lack of clarity within teams about who was responsible for entering these data and what training was required. Therefore, we chose to add another "workflow meeting" between key program staff and our team to help programs identify who was responsible for which surveys, who needed training, and how programs could monitor survey completion. We added these workflow meetings to our formal training protocol and made the support materials available to all sites who joined after the piloting phase.

Furthermore, our team worked with programs beyond the piloting phase to ensure that we continued to incorporate individual feedback and offer continuous support, which is key to successful adoption and can improve engagement [46]. After the piloting of the beta version of Beehive concluded, we continued to make development changes to meet users' needs, such as further design changes to the admin dashboard, widening survey completion windows, adjusting and eventually allowing customization of the frequency and timing of web link notifications, allowing the EULA to be completed before survey baseline date, simplifying registration fields, adding a survey status page, adding additional survey visualizations, adding a workflow for providers to enter data collected outside of Beehive, and prioritizing the order of additional languages in the app based on active need in the participating clinics. This iterative approach in response to user feedback is consistent with the development process of other eHealth apps [47].

This study highlights how critical it is for programs using a continuous improvement approach, such as UCD, to budget appropriately for ongoing development needs and staff time for ongoing support. As long as an app is in use and collecting data from real users, there should be a plan for ongoing project management and app development to address feedback from users, improve engagement and useability, and respond to changing needs. Implementing UCD from the outset allowed our team to be aware of and address user concerns before investing valuable time and resources in initial development and implementation. Focusing on workflow during the storyboard and alpha phases of app and continuing this collaborative relationship throughout the implementation phase resulted in an app that represents the interests and needs of users.

During the piloting, we observed that survey completion rates varied among different types of users. This variance may be partially explained by our training approach during piloting (see the Results section). As we continue to collect data after the piloting phase, we can evaluate if this trend continues beyond

the initial onboarding period throughout multiple years of data collection. These varied results may also reflect the challenges of implementing MBC, with or without an eHealth app, such as the training burden and limited time to conduct new duties associated with eHealth implementation [24-26,31]. When implementing outcomes data collection in these settings, it may be critical to gather only the minimum required data from EPI program providers (eg, diagnosis) and rely on service-user self-report measures whenever possible. Future analyses will examine the relationship between characteristics of EPI program providers (eg, degree and years licensed) and completion rates of clinician-entered data. Future work, including barriers and facilitators interviews with users after they gain more experience using Beehive, will be used to prioritize the needs and perspectives of our users in the ongoing development of Beehive and to better understand the reasons why users do or do not engage with the app [48].

Limitations

The COVID-19 pandemic introduced multiple challenges for our study, which may have reduced the breadth and diversity of participation in various phases of the project. We offered workshops over remote teleconferencing instead of in-person, which may have excluded individuals who are less comfortable with using technology. This may have disproportionately impacted on the recruitment of service users and support persons for participation in workgroups, as our participant numbers were lower than previous studies where we were conducting in-person research [31]. To reduce bias that may have resulted from this imbalance, we chose to prioritize the feedback of service users and support persons if there was conflicting feedback between these participants and EPI program provider participants. In addition, the beta and full versions of Beehive have been introduced to all service users in participating programs, regardless of their comfort with technology, and this has allowed us to incorporate informal feedback from these individuals as we have continued to make improvements to Beehive.

Much of our data in UCD workshops were gathered at the start of the COVID-19 pandemic, before anyone had experienced the long-term shift in daily practices brought on by the increased use of telehealth and remote working. We sought feedback on an app that was intended for in-person use and received feedback based on participants' experience of using in-person services. This highlights the importance of planning multiple opportunities for soliciting and incorporating feedback from sites so that apps can be responsive to changing environments.

Our workshops and piloting were only conducted in English. To serve the diverse population of California, Beehive needs to be both translated and adapted, a process known as localization [49], into at least 15 languages. Since Beehive's launch, we have localized into 7 additional languages. We continue to solicit feedback from users, including those whose primary language is not English, to inform the ongoing development of Beehive, and we will continue localize this app into additional languages.

While we used our prior knowledge of mental health development in the development of Beehive [11,12], we did not use a structured analysis approach for the feedback obtained during workshops due to time constraints imposed by project deliverables. To reduce the impact of subjective biases, the researchers who conducted each group debriefed afterward to review the notes, and recordings were referenced if notes were unclear or vague. In addition, all decisions about how to incorporate feedback from these notes into app development were made collaboratively by authors KEB, LMT, TAN, and VLT. Future work in this area will benefit from organized approaches to data collection and formal qualitative analysis [50,51].

Conclusions

Working with community partners to co-design an eHealth app for use in community EPI programs helped us to anticipate and resolve barriers earlier in the app development and implementation pipeline. On the basis of our observation and the data, there appeared to be high levels of engagement with Beehive. This resulted in feedback and continued design improvements which allowed our team to be better poised to launch Beehive across the EPI-CAL LHCN. Variance in survey completion rates among respondent types suggests that support persons and EPI program providers especially may need additional support.

Acknowledgments

The authors would like to thank all service users, support persons, and EPI program providers who participated in this study, Binda Mangat of Quorum Technologies, and its team of professionals for supporting the development of Beehive, and the Early Psychosis Intervention Network of California learning health care network.

This research made was supported by One Mind, the National Institute of Mental Health under award R01MH120555-01, and the following California counties: Los Angeles, Napa, Orange, San Diego, Solano, Stanislaus, Lake, Nevada, Mono, Colusa, and Sonoma. The content is solely the responsibility of the authors and does not necessarily represent the official views of any sponsors of the research.

Data Availability

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

Authors' Contributions

KEB conceptualized and designed the study, created the focus group guides, conceptualized and designed storyboards, recruited participants, collected data, trained pilot sites, audited and cleaned data, analyzed data, drafted the manuscript, and reviewed and revised the manuscript for important intellectual content. VLT conceptualized and designed the study, created the focus group guides, conceptualized and designed storyboards, collected data, trained pilot sites, drafted the manuscript, and reviewed and revised the manuscript for important intellectual content. KMP audited and cleaned data, drafted the manuscript, and reviewed and revised the manuscript for important intellectual content. LMT conceptualized and designed the study, created the focus group guides, conceptualized and designed storyboards, collected data, trained pilot sites, and reviewed and revised the manuscript for important intellectual content. SE conceptualized and designed the study, created the focus group guides, collected data, trained pilot sites, and reviewed and revised the manuscript for important intellectual content. MS, ABW, and AJP conceptualized and designed the study and reviewed the manuscript for important intellectual content. LS collected data, trained pilot sites, and reviewed and revised the manuscript for important intellectual content. CKH trained pilot sites, audited and cleaned data, and reviewed and revised the manuscript for important intellectual content. VEP, APM, and MK-W trained pilot sites and reviewed and revised the manuscript for important intellectual content. CM, MJM, NS, KLHN, and YZ cleaned data and reviewed and revised the manuscript for important intellectual content. TAN obtained funding, conceptualized and designed the study, created the focus group guides, conceptualized and designed storyboards, collected data, reviewed and revised the manuscript for important intellectual content.

Conflicts of Interest

KEB consulted with ChatOwl Inc after data collection and before submission. LMT was employed by ChatOwl Inc, a digital mental health company after data collection and before submission. LMT was a founder and owner of shares in Safari Health, Inc. during project implementation, data collection, and write up (no longer true at time of writing) and is an employee and shareholder at Kooth LLC at time of publication. TAN is a cofounder and shareholder in Safari Health, Inc.

Multimedia Appendix 1
The Early Psychosis Intervention Network of California outcomes collected in Beehive.
[DOCX File , 46 KB - humanfactors_v12i1e65889_app1.docx]

References

1. van Os J, Hanssen M, Bijl RV, Vollebergh W. Prevalence of psychotic disorder and community level of psychotic symptoms: an urban-rural comparison. Arch Gen Psychiatry 2001 Jul;58(7):663-668. [doi: 10.1001/archpsyc.58.7.663] [Medline: 11448373]
2. Menchaca A, Pratt B, Jensen E, Jones N. Examining the racial and ethnic diversity of adults and children. United States Census Bureau. 2023 May 22. URL: https://tinyurl.com/3wbpwc8t [accessed 2024-08-16]
3. Niendam TA, Sardo A, Savill M, Patel P, Xing G, Loewy RL, et al. The rise of early psychosis care in California: an overview of community and university-based services. Psychiatr Serv 2019 Jun 01;70(6):480-487. [doi: 10.1176/appi.ps.201800394] [Medline: 30890048]
4. Tryon VL, Nye KE, Savill M, Loewy R, Miles MJ, Tully LM, et al. The California collaborative network to promote data driven care and improve outcomes in early psychosis (EPI-CAL) project: rationale, background, design and methodology. BMC Psychiatry 2024 Nov 14;24(1):800 [FREE Full text] [doi: 10.1186/s12888-024-06245-6] [Medline: 39543502]
5. Heinssen RK, Azrin ST. A national learning health experiment in early psychosis research and care. Psychiatr Serv 2022 Sep 01;73(9):962-964 [FREE Full text] [doi: 10.1176/appi.ps.20220153] [Medline: 35895842]
6. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. JAMA Psychiatry 2019 Mar 01;76(3):324-335 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3329] [Medline: 30566197]
7. Firth J, Cotter J, Torous J, Bucci S, Firth JA, Yung AR. Mobile phone ownership and endorsement of "mHealth" among people with psychosis: a meta-analysis of cross-sectional studies. Schizophr Bull 2016 Mar;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
8. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
9. Meyer N, Kerz M, Folarin A, Joyce DW, Jackson R, Karr C, et al. Capturing rest-activity profiles in schizophrenia using wearable and mobile technologies: development, implementation, feasibility, and acceptability of a remote monitoring platform. JMIR Mhealth Uhealth 2018 Oct 30;6(10):e188 [FREE Full text] [doi: 10.2196/mhealth.8292] [Medline: 30377146]
10. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jan 21;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]

11. Kumar D, Tully LM, Iosif AM, Zakskorn LN, Nye KE, Zia A, et al. A mobile health platform for clinical monitoring in early psychosis: implementation in community-based outpatient early psychosis care. JMIR Ment Health 2018 Feb 27;5(1):e15 [FREE Full text] [doi: 10.2196/mental.8551] [Medline: 29487044]
12. Niendam TA, Tully LM, Iosif AM, Kumar D, Nye KE, Denton JC, et al. Enhancing early psychosis treatment using smartphone technology: a longitudinal feasibility and validity study. J Psychiatr Res 2018 Jan;96:239-246. [doi: 10.1016/j.jpsychires.2017.10.017] [Medline: 29126059]
13. Institute of Medicine, Committee on the Learning Health Care System in America, McGinnis JM, Stuckhardt L, Smith M, Saunders R. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America. Washington, DC: National Academies Press; 2013.
14. Unützer J, Katon W, Callahan CM, Williams JWJ, Hunkeler E, Harpole L, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA 2002 Dec 11;288(22):2836-2845. [doi: 10.1001/jama.288.22.2836] [Medline: 12472325]
15. Katon WJ, Lin EH, Von Korff M, Ciechanowski P, Ludman EJ, Young B, et al. Collaborative care for patients with depression and chronic illnesses. N Engl J Med 2010 Dec 30;363(27):2611-2620 [FREE Full text] [doi: 10.1056/NEJMoa1003955] [Medline: 21190455]
16. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. Cogn Behav Pract 2015 Feb;22(1):49-59 [FREE Full text] [doi: 10.1016/j.cbpra.2014.01.010] [Medline: 27330267]
17. Heinssen RK, Goldstein AB, Azrin ST. Evidence-based treatments for first episode psychosis: components of coordinated specialty care. National Institutes Mental Health. 2014 Apr 14. URL: https://www.nimh.nih.gov/sites/default/files/documents/health/topics/schizophrenia/raise/evidence-based-treatments-for-first-episode-psychosis.pdf [accessed 2025-03-31]
18. Garland AF, Kruse M, Aarons GA. Clinicians and outcome measurement: what's the use? J Behav Health Serv Res 2003;30(4):393-405. [doi: 10.1007/BF02287427] [Medline: 14593663]
19. Bickman L. A measurement feedback system (MFS) is necessary to improve mental health outcomes. J Am Acad Child Adolesc Psychiatry 2008 Oct;47(10):1114-1119 [FREE Full text] [doi: 10.1097/CHI.0b013e3181825af8] [Medline: 20566188]
20. Trivedi MH, Daly EJ. Measurement-based care for refractory depression: a clinical decision support model for clinical research and practice. Drug Alcohol Depend 2007 May;88 Suppl 2(Suppl 2):S61-S71 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.01.007] [Medline: 17320312]
21. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
22. Treisman GJ, Jayaram G, Margolis RL, Pearlson GD, Schmidt CW, Mihelish GL, et al. Perspectives on the use of eHealth in the management of patients with schizophrenia. J Nerv Ment Dis 2016 Aug;204(8):620-629 [FREE Full text] [doi: 10.1097/NMD.0000000000000471] [Medline: 26828911]
23. Ossebaard HC, van Gemert-Pijnen L. eHealth and quality in health care: implementation time. Int J Qual Health Care 2016 Jun;28(3):415-419. [doi: 10.1093/intqhc/mzw032] [Medline: 27029590]
24. Hatfield DR, Ogles BM. Why some clinicians use outcome measures and others do not. Adm Policy Ment Health 2007 May;34(3):283-291. [doi: 10.1007/s10488-006-0110-y] [Medline: 17211715]
25. Zimmerman M, McGlinchey JB. Why don't psychiatrists use scales to measure outcome when treating depressed patients? J Clin Psychiatry 2008 Dec;69(12):1916-1919. [doi: 10.4088/jcp.v69n1209] [Medline: 19192467]
26. Jensen-Doss A, Haimes EM, Smith AM, Lyon AR, Lewis CC, Stanick CF, et al. Monitoring treatment progress and providing feedback is viewed favorably but rarely used in practice. Adm Policy Ment Health 2018 Jan;45(1):48-61 [FREE Full text] [doi: 10.1007/s10488-016-0763-0] [Medline: 27631610]
27. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985;28(3):300-311. [doi: 10.1145/3166.3170]
28. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-computer Interaction. Mahwah, NJ: Lawrence Erlbaum Associates; 1986.
29. Gulliksen J, Göransson B, Boivie I, Blomkvist S, Persson J, Cajander Å. Key principles for user-centred systems design. Behav Inform Technol 2003 Nov;22(6):397-409. [doi: 10.1080/01449290310001624329]
30. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
31. Savill M, Banks LM, Tryon VL, Ereshefsky S, Nye KE, Botello RM, et al. Exploring data collection priorities of community partners in early psychosis care. Psychiatr Serv 2024 Sep 01;75(9):854-862. [doi: 10.1176/appi.ps.20230455] [Medline: 38595117]
32. Tully LM, Nye KE, Ereshefsky S, Tryon VL, Hakusui CK, Savill M, et al. Incorporating community partner perspectives on eHealth technology data sharing practices for the California early psychosis intervention network: qualitative focus group study with a user-centered design approach. JMIR Hum Factors 2023 Nov 14;10:e44194 [FREE Full text] [doi: 10.2196/44194] [Medline: 37962921]

33. Endmann A, Keßner D. User journey mapping – a method in user experience design. i-com 2016;15(1):105-110. [doi: 10.1515/icom-2016-0010]
34. Newman MW, Landay JA. Sitemaps, storyboards, and specifications: a sketch of Web site design practice. In: Proceedings of the 3rd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques. 2000 Presented at: DIS '00; August 17-19, 2000; New York, NY. [doi: 10.1145/347642.347758]
35. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015 Jul 10;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
36. Core Assessment Battery (CAB). Early Psychosis Intervention Network. URL: https://nationalepinet.org/core-assessment-battery-cab/ [accessed 2024-08-16]
37. Boothroyd RA, Chen HJ. The psychometric properties of the Colorado Symptom Index. Adm Policy Ment Health 2008 Sep;35(5):370-378. [doi: 10.1007/s10488-008-0179-6] [Medline: 18561020]
38. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017;8:72 [FREE Full text] [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]
39. van Bruinessen IR, van Weel-Baumgarten EM, Snippe HW, Gouw H, Zijlstra JM, van Dulmen S. Active patient participation in the development of an online intervention. JMIR Res Protoc 2014 Nov 06;3(4):e59 [FREE Full text] [doi: 10.2196/resprot.3695] [Medline: 25379679]
40. van Hierden Y, Dietrich T, Rundle-Thiele S. Designing an eHealth well-being program: a participatory design approach. Int J Environ Res Public Health 2021 Jul 06;18(14):7250 [FREE Full text] [doi: 10.3390/ijerph18147250] [Medline: 34299700]
41. Honary M, Fisher NR, McNaney R, Lobban F. A web-based intervention for relatives of people experiencing psychosis or bipolar disorder: design study using a user-centered approach. JMIR Ment Health 2018 Dec 07;5(4):e11473 [FREE Full text] [doi: 10.2196/11473] [Medline: 30530457]
42. Mak C, Whittingham K, Cunnington R, Boyd RN. Efficacy of mindfulness-based interventions for attention and executive function in children and adolescents—a systematic review. Mindfulness 2017 Jul 26;9(1):59-78. [doi: 10.1007/s12671-017-0770-6]
43. Kajitani K, Higashijima I, Kaneko K, Matsushita T, Fukumori H, Kim D. Short-term effect of a smartphone application on the mental health of university students: a pilot study using a user-centered design self-monitoring application for mental health. PLoS One 2020;15(9):e0239592 [FREE Full text] [doi: 10.1371/journal.pone.0239592] [Medline: 32976515]
44. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
45. The value of beehive. YouTube. URL: https://www.youtube.com/watch?v=-7MDPRLU1BM&feature=youtu.be [accessed 2024-08-29]
46. Glomann L, Hager V, Lukas CA, Berking M. Patient-centered design of an e-mental health app. In: Joint Proceedings of the AHFE 2018 International Conference on Human Factors in Artificial Intelligence and Social Computing, Software and Systems Engineering, The Human Side of Service Engineering and Human Factors in Energy. 2018 Presented at: AHFE 2018; July 21-25, 2018; Orlando, FL. [doi: 10.1007/978-3-319-94229-2_25]
47. Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders SM. Methods for human-centered eHealth development: narrative scoping review. J Med Internet Res 2022 Jan 27;24(1):e31858 [FREE Full text] [doi: 10.2196/31858] [Medline: 35084359]
48. Ereshefsky S, Gemignani R, Savill M, Sanford KC, Banks LM, Tryon VL, et al. A mixed-methods study exploring the benefits, drawbacks, and utilization of data in care: findings from the EPI-CAL early psychosis learning health care network. Schizophr Res 2025 Feb;276:157-166 [FREE Full text] [doi: 10.1016/j.schres.2025.01.018] [Medline: 39892249]
49. Schäler R. Localization and translation. In: Handbook of Translation Studies Online. Amsterdam, The Netherlands: John Benjamins Publishing Company; 2010:209-214.
50. Molina-Recio G, Molina-Luque R, Jiménez-García AM, Ventura-Puertos PE, Hernández-Reyes A, Romero-Saldaña M. Proposal for the user-centered design approach for health apps based on successful experiences: integrative review. JMIR Mhealth Uhealth 2020 Apr 22;8(4):e14376 [FREE Full text] [doi: 10.2196/14376] [Medline: 32319965]
51. Ozkaynak M, Sircar CM, Frye O, Valdez RS. A systematic review of design workshops for health information technologies. Informatics 2021 May 14;8(2):34. [doi: 10.3390/informatics8020034]

Abbreviations

EPI: early psychosis intervention
EPI-CAL: Early Psychosis Intervention Network of California
EULA: end user license agreement
HIPAA: Health Insurance Portability and Accountability Act

LHCN: learning health care network
MBC: measurement-based care
MCSI: Modified Colorado Symptom Index
UCD: user-centered design

Edited by C Jacob; submitted 04.11.24; peer-reviewed by J D'Arcey; comments to author 06.12.24; revised version received 19.12.24; accepted 12.03.25; published 09.04.25.

Please cite as:
Burch KE, Tryon VL, Pierce KM, Tully LM, Ereshefsky S, Savill M, Smith L, Wilcox AB, Hakusui CK, Padilla VE, McNamara AP, Kado-Walton M, Padovani AJ, Miller C, Miles MJ, Sharma N, Nguyen KLH, Zhang Y, Niendam TA
Co-Designing a Web-Based and Tablet App to Evaluate Clinical Outcomes of Early Psychosis Service Users in a Learning Health Care Network: User-Centered Design Workshop and Pilot Study
JMIR Hum Factors 2025;12:e65889
URL: https://humanfactors.jmir.org/2025/1/e65889
doi:10.2196/65889
PMID:

On the Necessity of Multidisciplinarity in the Development of at-Home Health Monitoring Platforms for Older Adults: Systematic Review

Chris Lochhead, BSCE, MSc; Robert B Fisher, BSc, MSc, PhD

Advanced Care Research Centre, University of Edinburgh, Old College, South Bridge, Edinburgh, United Kingdom

Corresponding Author:
Chris Lochhead, BSCE, MSc
Advanced Care Research Centre, University of Edinburgh, Old College, South Bridge, Edinburgh, United Kingdom

Abstract

Background: The growth of aging populations globally has increased the demand for new models of care. At-home, computerized health care monitoring is a growing paradigm, which explores the possibility of reducing workloads, lowering the demand for resource-intensive secondary care, and providing more precise and personalized care. Despite the potential societal benefit of autonomous monitoring systems when implemented properly, uptake in health care institutions is slow, and a great volume of research across disciplines encounters similar common barriers to real-world implementation.

Objective: The goal of this systematic review was to construct an evaluation framework that can assess research in terms of how well it addresses already identified barriers to application and then use that framework to analyze the literature across disciplines and identify trends between multidisciplinarity and the likelihood of research being developed robustly.

Methods: This paper introduces a scoring framework for evaluating how well individual pieces of research address key development considerations using 10 identified common barriers to uptake found during meta-review from different disciplines across the domain of health care monitoring. A scoping review is then conducted using this framework to identify the impact that multidisciplinarity involvement has on the effective development of new monitoring technologies. Specifically, we use this framework to measure the relationship between the use of multidisciplinarity in research and the likelihood that a piece of research will be developed in a way that gives it genuine practical application.

Results: We show that viewpoints of multidisciplinarity; namely across computer science and medicine alongside public and patient involvement (PPI) have a significant positive impact in addressing commonly encountered barriers to application research and development according to the evaluation criteria. Using our evaluation metric, multidisciplinary teams score on average 54.3% compared with 35% for teams made up of medical experts and social scientists, and 2.68 for technical-based teams, encompassing computer science and engineering. Also identified is the significant effect that involving either caregivers or end users in the research in a co-design or PPI-based capacity has on the evaluation score (29.3% without any input and between 48.3% and 36.7% for end user or caregiver input respectively, on average).

Conclusions: This review recommends that, to limit the volume of novel research arbitrarily re-encountering the same issues in the limitations of their work and hence improve the efficiency and effectiveness of research, multidisciplinarity should be promoted as a priority to accelerate the rate of advancement in this field and encourage the development of more technology in this domain that can be of tangible societal benefit.

(JMIR Hum Factors 2025;12:e59458) doi:10.2196/59458

KEYWORDS

multi-disciplinarity; gait assessment; machine learning; at-home health monitoring; older adults; elderly; artificial intelligence; AI; gait; development; health monitoring; monitoring; systematic review; monitoring system; barriers; caregiver; efficiency; effectiveness

Introduction

Background

With the general increase in global life expectancy and the ever-increasing ratio of retirement-age to working-age people, research in the scope of health and social care for older people is greatly increasing. As people live longer and the aging population grows, governments and international organizations are recognizing the need for a paradigm shift in the way we care for our older populations [1]. Current methods of care typically include a social care system to look after the most vulnerable and frail older adults, but as the size of this population increases at a faster rate than the working population, these institutional

models of care at the country level become progressively less practical and more costly. In response to this growing crisis, one solution researchers have sought to apply is the use of novel health monitoring technologies to tackle the problem of labor shortages in health care systems and to improve productivity and efficiency in care [2-11]. This can be in direct terms, for example monitoring systems that assist care professionals carrying out their duties within care settings [12] and also indirectly with at-home monitoring devices designed to reduce the occurrence of injuries by, for example, assessing gait [13,14] and predicting fall events [3,15], with the goal of reducing the potential burden on secondary care and ultimately reducing the burden on the care-home sector and keeping people at home, healthy, and independent for longer [9,16,17].

Research Hypothesis and Goals

The goals of this work can be summarized by the following 2 research questions:

1. Can we construct an evaluation framework for at-home health monitoring research and justify that a positive score in said framework broadly correlates with a higher likelihood of the research being effective in real-world use?
2. Can we identify the trends if they exist, between the method of research (single-discipline, multidiscipline, and the use of PPI or co-design) and the increase or decrease in effectiveness of said research?

In the context of this paper, “effective research” can be defined as research that is more efficient by being less prone to making oversights already identified in the existing literature across disciplines. We assert that research that addresses the already existing issues in the literature (and thereby scoring highly on our proposed metric) leads to more effective research and ultimately leads to products and platforms more likely to be usable in real-life applications (refer to the Evaluation Framework Justification in the Methods section for evidence of this).

To answer the research questions, this work is split into 2 phases, a scoping phase and evaluation phase. In the scoping phase, an overview is provided of the various systematic reviews across disciplines in the field of at-home health care monitoring for older adults. In doing this, we synthesize an evaluation framework based on the consensus and concatenation of these studies to measure the degree to which individual pieces of research encounter similar common problems. By defining this framework, observations can be made about the degree to which these problems are acknowledged by different types of research teams operating with or without multidisciplinarity, which types of issues are likely to be identified by which types of research, and how stakeholder involvement can improve the likelihood of developing an effective product. To justify this evaluation criteria and to address the first research question, examples of existing technologies in commercial use are positively evaluated by this approach to assert the positive relationship between a high evaluation score and the generation of more effective research, considerate to existing identified issues with real-world use (refer to the Methods section). To our knowledge, there is currently no evaluation methodology designed to measure research application effectiveness, based on multidisciplinary metrics.

In the evaluation phase. a review of individual works in the application of computerized at-home health-monitoring systems is then presented, with these works evaluated against the metric developed in the scoping phase, and with the results analyzed to address the second research question (refer to the Results section). For this evaluation phase, a total of 350 papers were found from the IEEE Explore, PubMed, and ArXiv databases, of which 60 papers were ultimately used, with the inclusion constraints being that the papers had to be individual pieces of original research relating either to the development of technology associated with the various applications of at-home older adult health monitoring or a review into the effectiveness of existing at-home health monitoring applications. This is including, but not limited to, development and deployment of disease diagnosis and progression analysis, fall detection and prevention, lifestyle monitoring, vital-sign monitoring, and smart-home systems. Candidate papers must present methods that have the prospect of or are already actively being tested in an at-home environment in whole or in part.

For this evaluation phase, a total of 350 papers were found from the IEEE Explore, PubMed, and ArXiv databases, of which 60 papers were ultimately used, with the inclusion constraints being that the papers had to be individual pieces of original research relating either to the development of technology associated with the various applications of at-home older adult health monitoring or a review into the effectiveness of existing at-home health monitoring applications. This is including but not limited to the development and deployment of disease diagnosis and progression analysis, fall detection and prevention, lifestyle monitoring, vital-sign monitoring, and smart-home systems. Candidate papers must present methods that have the prospect of or are already actively being tested in an at-home environment in whole or in part.

The Methods section contains the scoping phase and provides an overview of the literature in health monitoring technologies, constructing from them a concise evaluation framework in the form of a 10-point checklist criteria for research application effectiveness, based on consensus drawn from these works. The Results section presents the evaluation phase, reviewing individual pieces of research across disciplines using this evaluation framework. The interaction between multidisciplinarity and stakeholder involvement through PPI and co-design and the effectiveness of research are also analyzed here. Finally, the Discussion section summarizes the findings of this research and makes recommendations for how future health monitoring research should be conducted in order to improve the effectiveness of research in this field.

Research Contributions

The contributions of this paper can be summarized as follows:
:

1. A review of the recent literature of at-home health care monitoring. This is achieved by providing a 2-stage review of the literature, with the first (scoping phase) being a meta-review of reviews in the relevant literature, and the

second (evaluation phase) being a review of individual works across multiple disciplines in the area of at-home health monitoring.

2. The introduction of a comprehensive evaluation framework for assessing the likely real-world effectiveness of health monitoring application research, based on 10 key issues consistently identified from analysis of 15 literature reviews in the field across disciplines.
3. The identification and demonstration of various links between the manner of research and the effectiveness of research. We find that multidisciplinarity has a consistently positive effect in this respect, reflected by higher scores on the evaluation framework. We also find a convincing increase in scores when PPI and co-design methods are used.

The goal of this research is to concretely illustrate the necessity of multidisciplinarity for the success of at-home medical technology development and deployment and provide a framework for future research to reference when assessing the usefulness of their own applications.

Methods

Scoping Phase: Overview of at-Home Health-Monitoring in Older Adults

To identify the most common barriers to uptake in this field, an overview of 12 systematic and other review papers from 2014 onwards was conducted to identify and collate the common issues within some of the common subdisciplines of at-home health care monitoring: encompassing any autonomous monitoring methods used in a domestic context such as camera-based applications, wearable sensors, remote sensors, and ensemble smart-home systems. Table 1 provides a concise summary of the main reviews investigated, and the presence of each of the commonly identified barriers among them. Broken down in Table 1 are the makeup of the teams involved in the research, multidisciplinary, technical, or application-based. In the context of this work, “Technical” concerns research conducted by engineering or computer science teams, with “Application” making up all other types of research team, but predominantly teams in the fields of medicine and social science. In total, 10 core barriers are identified across these reviews and synthesized to be inclusive of all the barriers identified in all the systematic reviews evaluated in this section. An exact breakdown of the presence or absence of the 10 points in each review is available in the supplementary materials (refer Table S1 in Multimedia Appendix 1).

Table . Overview of systematic reviews and their evaluation score according to the evaluation metric.

Paper title and reference	Description	Research team	Score (%)
Are Active and Assisted Living applications addressing the main acceptance concerns of their beneficiaries? [18]	Overview of opinions of older adults regarding concerns with ambient assisted living technologies.	Multidisciplinary	50
A critical review of smart residential environments for older adults with a focus on pleasurable experience [5]	Review of older adults in focus groups for a variety of smart home applications.	Multidisciplinary	30
Smart homes and home health monitoring technologies for older adults: A systematic review [6]	Investigation of the abilities of various smart-home technologies, alongside feasibility and technical limitations.	Application	70
Health Monitoring Using Smart Home Technologies: Scoping Review [8]	Review of smart home environments, specifically on existing study design limitations alongside technical limitations.	Application	50
Older persons have ambivalent feelings about the use of monitoring technologies [2]	Series of 5 focus groups of older adults attempting to build consensus on the concerns of smart home implementations.	Application	60
Older adults' perceptions of technologies aimed at falls prevention [3]	Systematic review of focus group-based studies to build consensus on the main factors hindering at home uptake.	Application	40
Unobtrusive sensing and wearable devices for health informatics [11]	Overview of 4 main sensor-based monitoring technologies including relative benefits and drawbacks.	Technical	40
Wearable sensors for remote health monitoring [7]	Overview of specifically wearable health monitoring technologies, mostly centered on technical benefits and limitations regarding data.	Technical	40
Unobtrusive health monitoring in private spaces: The smart home [10]	Overview of smart home. Applications with a focus on perceived "unobtrusive" applications.	Technical	20
Detection and assessment of Parkinson's disease based on gait analysis: A survey [14]	Overview of gait assessment monitoring for age-related disease detection using a variety of classical and ML[a]-powered monitoring technology.	Technical	0
Remote patient monitoring using AI[b] [9]	Overview of classical and deep learning-based AI applications in primarily at-home patient monitoring.	Technical	20
Factors Determining the Success and Failure of eHealth Interventions [16]	Overview of smart home. Applications with a focus on discovering the key factors behind the success or failure of med-tech applications.	Multidisciplinary	70

[a]ML: machine learning.

[b]AI: artificial intelligence.

"Research Team" denotes whether the teams conducting the research were technical or medical-based. See further in this section for a breakdown of the 10 common identified points and refer to the supplementary material for a specific breakdown of which barriers are present in each. The score value was computed by aggregating which of the 10 features were addressed in each review. Beyond these key reviews from which the evaluation framework was constructed, there are numerous other reviews and surveys whose findings are inclusive of the framework outlined at the end of this section (refer to Table S2 in Multimedia Appendix 1 for a full breakdown of the 13 reviews). Only the core 12 are included here for brevity, with these 12 being selected as they encompass all of the core at-home health care methodologies, include examples of all 10

key points that make up the framework, and represent a roughly equal variety of research teams, namely multidisciplinary teams (n=3), computer science (n=3) engineering (n=2), medical (n=2), and social science (n=2).

Table . Scoring of 2 current at-home monitoring projects being developed with the intention of widespread use. The factors are numbered corresponding to the attribute list given in the attribute list above. Y or N indicates "yes" or "no" as to the presence of the factors.

Project Name	Description	Factors (Y or N)									
		1	2	3	4	5	6	7	8	9	10
SPHERE[a] [19-21]	Smart-home system for behavior monitoring, under development for several years and the feature of multiple research studies.	Y[b]	Y	Y	Y	Y	Y	Y	Y	N[c]	Y
HALLEY[d] [22]	Internet of things-based Smart-home development project currently in commercial development.	Y	Y	Y	N	Y	Y	Y	N	N	Y

[a] SPHERE is a multi-year smart-home development project based in England which has been involved both in commercial smart-home production as well as research.

[b] Y: yes.

[c] N: no.

[d] HALLEYASSIST is a company based in Australia involved in developing multiple healthcare monitoring devices for a smart home context.

The study by Ghorayeb et al [23] presents a number of insights in their review of the literature regarding older adults using smart-homes, such as identifying solutions to the problems in most research applications. For example, they assert that gradual introduction of smart-home technology combined with the ability to “pause” it at will to provide “emotional release” is highly desired in end users for a more pleasant experience. They also find that more tech-literate people tend to have less concerns with the technology due to an improved understanding of the data being collected and the privacy risks involved, if any. One other concern identified was transmission of data and the insecurities associated with this, leaning to a user-preference that data should be handled manually instead of through remote connection. These findings are all further reinforced by the findings in [24], which similarly conduct interview sessions with 20 older adults regarding which aspects of health-monitoring systems concern them. These findings inform points 4 and 5 in the evaluation metric, defined in Methods section.

Regarding data access, the study by Robinson et al [25] finds that older adults could in some cases want control of data access to withhold the data from certain parties, namely their family and friends, for fear of burdening them as well as the parallel desire some express that they do not want to be micromanaged by their loved ones or health care providers, informing points 6 and 8 of the framework. Cost was also an identified challenge (informing point 3), for example focus group attendees in [2] were concerned by the cost of long-term use of monitoring devices, with the implication being that they would have to buy them outright.

The study by Hawley-Hague et al [3] discovers that when surveyed, there is a broad perception that older people view new advancements in smart-home related technology as good but “unnecessary for them,” as they don’t perceive themselves as being “unhealthy enough” to merit using what they see as a drastic action toward greater care. Mann et al [26] found in their study of 661 older people that a slim majority (56.3%) perceived smart-home technology as not being of use to them specifically.

Ghorayeb et al [23] identify a series of factors through their investigation with older participants which are crucial to account for when considering the integration of new technology into real life. These factors range from societal stigma to technical reliability, covering points 1, 2, 3, and 6 in the evaluation framework. Research in studies by Boström et al and by Boström et al [2] and Liu et al [6] also point to the issues surrounding the implication of autonomous monitoring being a decline in human contact, addressed by point 7. In the systematic review conducted by Ghorayeb et al [23], observations were made regarding the progression of smart-home technologies. Regarding patient acceptability, they note that less than half of the papers reviewed take into consideration the acceptability of their technology where privacy is concerned: with the focus of the paper instead being explicitly about the functions of the novel technology. Some steps taken toward improving privacy in certain papers included encrypting collected data [27] and locking data access behind authorization [28].

Regarding health care professionals, their main concerns with smart-home technology and health care monitoring concern the feasibility of use. Unlike patient acceptability, acceptability by health care professionals is underresearched, where many papers will either focus on the technology being developed and not address it in application or they will focus only on the opinions of the end users and how they will use and accept the technology [29]. This finding is echoed in the lack of surveys found in the scoping review in the Results section, where the focus is on health care professionals rather than patients. Of the 60 papers in the scoping review, only 7 made any explicit use of external medical caregivers in their research development.

Evaluation Framework Criteria

Across the review papers both in Table 1 and the other reviews referenced in this section, the following list of 10 factors were collated to encompass all the common considerations identified when designing and implementing novel monitoring technology for use with older adults in a domestic environment. They are segmented by category of concern (technical, application, or multidisciplinary), where the technical are purely engineering or computer-science implementation concerns and application are concerns involving human interaction with the technology and any adjacent concerns important to medical and social science-based disciplines. The framework is as follows (Textbox 1).

Textbox 1. The evaluation framework

- Usability-technical: concerning the use of the technology and how feasible it is to be used by caregivers, health care professionals, and end users.
- Accessibility-technical: this concerns the ease of use by laypeople of the computer technology and the barriers for entry in terms of effectively using the technology.
- Reliability-technical: covers issues relating to the long-term viability of the application, for example, is it expensive, does it require upkeep, charging or maintenance.
- Control-multidisciplinary: concerns the level of access both patients and health care staff should have to the application and the data it records.
- Privacy-multidisciplinary: constitutes issues relating to the intrusiveness of the data being collected, and the manner in which it is collected.
- Stigma-application: anything relating to the societal pressures and negative connotations people may feel by using monitoring technologies for the purposes of personal health.
- Lack of human response-application: concerns the issue of the perception that increased autonomous monitoring would result in a decrease in person-to-person interaction as a result.
- Burden to others-application: regarding the perception of older adults that additional at-home monitoring is representative of applying additional pressure on their caregivers.
- Lack of perceived need-application: concerns a commonly identified phenomenon in the literature that people have a tendency to think a technology is useful but unnecessary for them personally.
- Affordability-application: this concerns the cost both on a personal and institutional level to implement solution applications in real life.

Evaluation Framework Justification

To demonstrate the descriptive power this evaluation system has to the likelihood of success in application, Table 2 provides an overview of papers concerning 2 technologies being developed for commercial use (Sphere [21] and HalleyAssist [22]), and the degree to which the prior research and development of these systems adhere to the criteria in the evaluation metric, with scores being calculated based on all existing literature regarding each technology, as opposed to individual articles. Both have demonstrated a commitment to addressing barriers relating to both technical and human elements, addressing 90% and 70% of the evaluation metrics respectively (the average score across the scoping review in the Results section is 36.2% for comparison). They also both were developed by a multidisciplinary research team from computer science and signal processing to nursing, geriatric medicine and social care, with both systems now being in varying stages of commercialization. The relationship between the absence of certain metrics such as multidisciplinarity and the lack of consideration for common barriers to uptake is further concretely illustrated during the analysis at the beginning of the Results section. While these are only 2 data points and cannot be said to constitute a trend on their own, it can be asserted that they provide a strong positive indication that successful applications are likely to score highly on the introduced metric.

Results

Evaluation Phase: Scoping Review and Parameters

Figure 1 illustrates the selection process for papers considered for the scoping review. From the 3 databases, the following search term was used:

((healthcare monitoring) AND (older adult)) AND ((home) OR (at-home) OR (domestic)) AND ((obtrusive) OR (unobtrusive) OR (intrusive)) AND ((machine learning) OR (AI) OR (artificial intelligence))

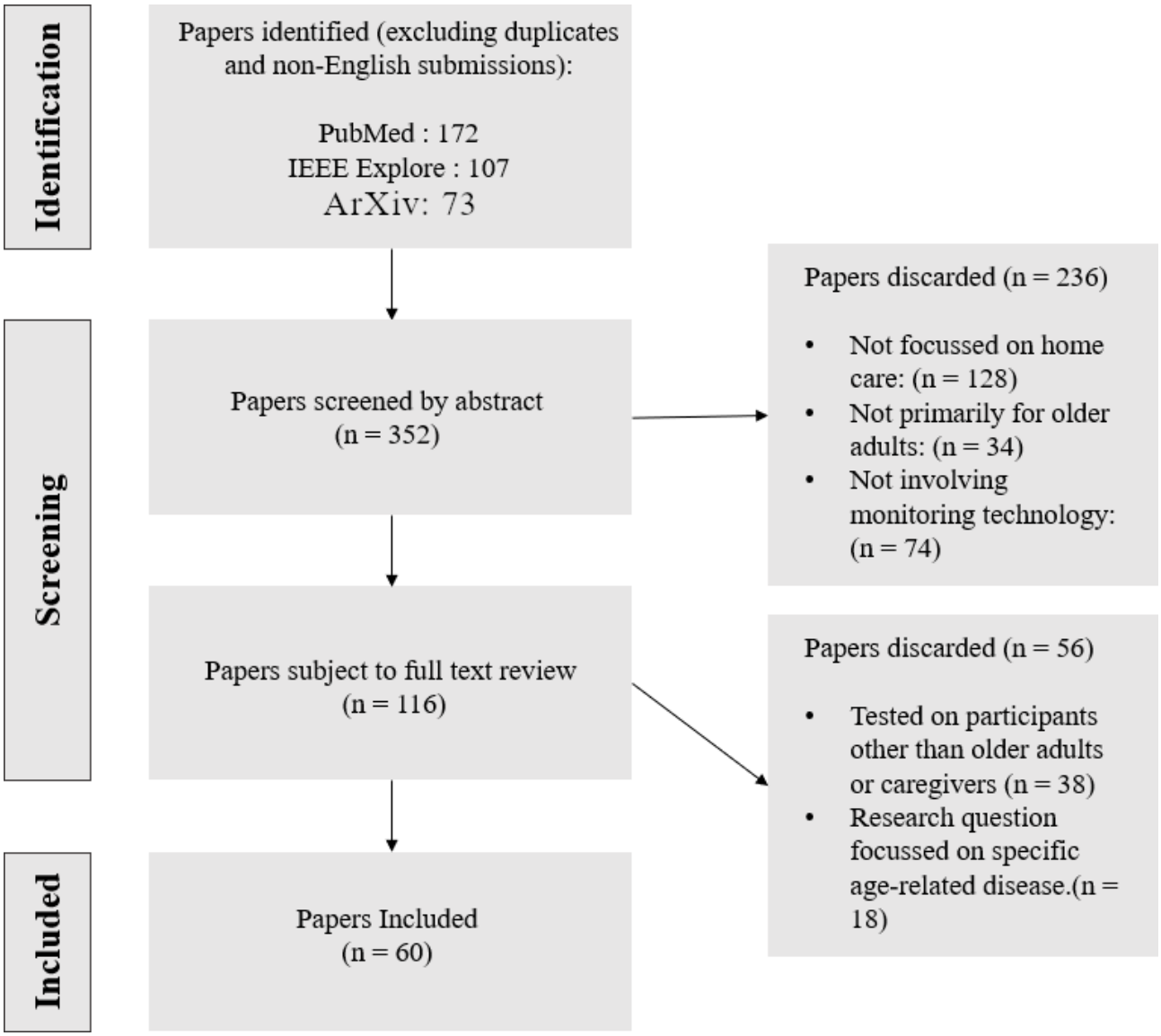
with the exception of the ArXiv database, where the

((obtrusive) OR (unobtrusive) OR (intrusive))

section was omitted for a larger range of initial texts. Language was restricted to English and the date was restricted to 2014 or later, with the search commencing on January 7-24, 2024. Excluded were papers that did not in whole or in part reference older adult monitoring as the purpose of their application or review, and papers where the application was specifically designed for care homes or hospitals. The initial search yielded a total of 352 papers. Paper title and abstract analysis excluded 236 papers, which resulted in 116 remaining, with the final count being 60 papers after a full-paper review resulting in 56 more deletions (Figure 1).

The following section is broken down by technical and application-based research, following the definitions for both established in the beginning of the Methods section. Papers defined as multidisciplinary are those in which the research team consist of at least 1 person from each category.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram: screening process illustration for the scoping review portion of this research.



Technical Research

The majority of research in this space relies either on traditional or neural network–based machine learning methods to make their systems effective and largely automatic. Researchers in [30] construct a machine learning model that can differentiate between regular, prefrail, and frail gait within a population of 50 older adults at an accuracy of 88.5% based on gait data

collected from wearable accelerometer signals. They rely mostly on “traditional” methods of machine learning: that is, machine learning methods that don’t use deep neural networks.

With vision-based applications, on the other hand, deep neural networks, especially convolutional networks designed specifically for processing image data, are frequently used. Researchers in the study by Lin et al [31] use a novel 3D convolutional neural network (CNN) to process gait silhouette images from the CASIA-B dataset [32] to achieve a person-identification accuracy of 97.6%. While the question of person identification through gait can be fairly trivially solved with modern CNN frameworks, the question of gait analysis, or assessing health requires a more nuanced approach and is still an active research area.

Researchers in [33] use a Spatiotemporal Graph Convolutional Network (ST-GCN) [34] framework to assess Parkinson disease severity as classified under the Unified Parkinson’s Disease Rating Scale [35]. The ST-GCN is an extension of the standard CNN model that works specifically on skeletal graph data as opposed to raw images, and works by processing input data with sequential attention to both the spatial and temporal dimensions of the input. As gait assessment is a far more complicated problem than simple person-identification or even traditional action recognition, they achieved 53% F_1-score using a dataset of 53 individuals with Parkinson disease.

Recently, in the study by Yin et al [36], the Spatio-temporal joint adjacency GCN (STJA-GCN) was developed that uses 3 input streams for joints, bones and velocities, a novel joint attention module to emphasize spatial data and a simplified skeletal graph input architecture to achieve state-of-the-art results (93.17% and 92.08%) on the domain of recognizing and classifying different types of abnormal gaits compared to other ST-GCN–based methodologies. This was tested on a pair of synthetic gait datasets introduced by the researchers, totalling 22 participants and 8600 instances of gait data collected from a total of 7 sensors across the 2 datasets.

Demonstrably, machine learning methods are extremely capable in multiple health care monitoring applications, necessitating the inclusion of highly technical fields in this type of research. The issue with the papers in this field, however, is that they score poorly on the proposed evaluation framework due to an overt focus on technical innovation at the expense of applicational concerns as well as scoring consistently lower than papers in the applicational and multidisciplinary categories on all but explicitly technical metrics (Figure 2). This analysis indicates a narrow focus almost purely on the effectiveness of the technology at the expense of an omission of application considerations, such as cost or perceived need by end users and medical stakeholders (refer to Figure 3 for a per-category breakdown). End users in this context refers both to patients and health care professionals who would be involved with the use of the technology (such as general practitioners, nurses, physiotherapists, occupational therapy and care workers). Stakeholders, on the other hand, refers to family members, friends or dependents of the patient using the technology. To take a more concrete example of this narrowed focus, research by Brunzini et al [13] is explicitly geared toward the development of a decision-support algorithm for health care professionals, complete with bespoke visual aids. However, there is no evidence of consideration of the requirements of health care professionals who would prospectively use such a system (in this case namely GPs, geriatricians, and nurses) being used in the design of the system itself. Similar issues were found in several studies where technical teams sought to develop assistive tools for caregivers or patients without direct involvement of prospective caregivers or end users, ranging from tools designed for doctors and nurses to carers, physiotherapists, and occupational therapists [37-41].

Figure 2. Illustration of evaluation performance per-evaluation category. Shown here is the impact that the use of multidisciplinarity has on the likelihood of each of the 10 metrics being taken into consideration.

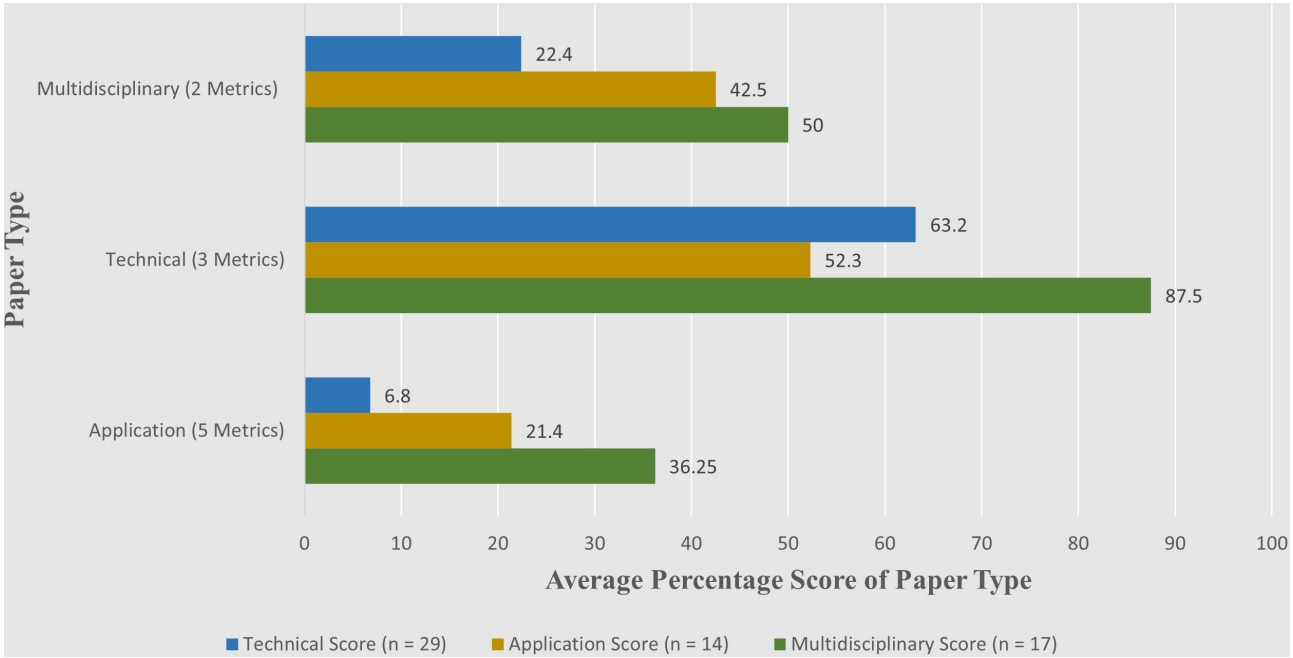
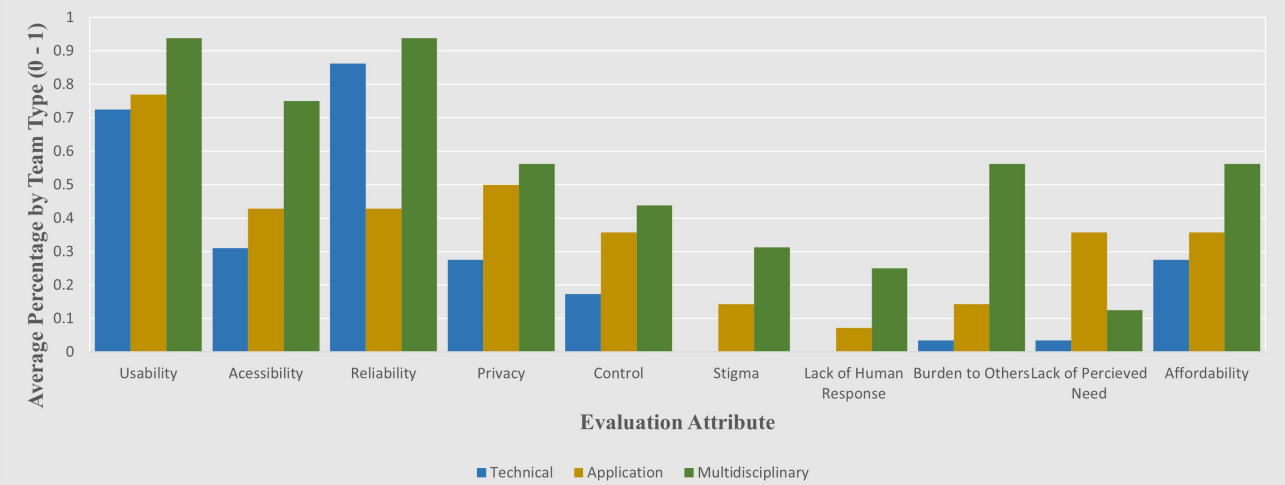


Figure 3. Breakdown of the average presence of each evaluation attribute by research team type.



Application and Multidisciplinary Research

Of the surveyed technical papers (n=29), they scored an average 6.8% on the application-centered metrics in the proposed evaluation framework (Figure 2). In contrast, the application papers (n=14) mainly authored by health care researchers scored 21.4% on application-based metrics. While this indicates that application focused papers typically address more concerns with real-life deployment of research solutions, there is also a notable trend that this focus on application comes at a slightly increased tendency to neglect technical aspects of research (with both technical and multidisciplinary papers outperforming application-based papers on technical metrics). A clear trend can be seen in multidisciplinary papers consistently outperforming technical or application-based papers, even on the metrics specific to their own discipline (Figure 3).

As indicated in the second research question in the Introduction, we speculate that there exists a relationship between the inclusion of different stakeholders in PPI and co-design with the effectiveness of the resultant applications and research. As indicated by Table 1, this type of inclusion is usually implemented using focus groups. In this section, we highlight the typically moderate-to-high scores where this stakeholder input is prioritised, whilst also outlining other areas of shortcomings that follow from lack of multidisciplinarity in other, mainly technical domains.

In the study by Mireles et al [42], a total of 75 health care professionals (mostly nurses, n=47) are interviewed as a means of investigating the effectiveness of 3 different eHealth technologies for at-home monitoring of older adults with chronic conditions. Scoring 40%, this research focuses on the implications of e-health solutions as they affect end users, namely stakeholders, with less consideration for the technical implications of their recommendations. Research in the study by Bjornsdottir and Ceci [43] is similar, conducting qualitative interviews with both health care professionals and also patients. By including the patients and getting their insight, this research scores higher (60%) but suffers on the same technical metrics, highlighting the shortcomings in the systems they are evaluating without explicit consideration of the technical implications of their conclusions. In [44], 5 focus groups were conducted with both older adults and caregivers; namely family and care staff, to help synthesize a consensus on the design of future at-home sensor-based systems, with a focus on acceptability, respect for privacy and how to best provide control of care to the end users themselves. Scoring 50%, the authors neglect aspects such as system reliability at the expense of greater privacy, and other application-driven barriers such as societal stigma and the threat of greater isolation resulting in the deployment of remote monitoring technology. Expanding on this focus-group centered research [45], conduct a series of interviews with both formal and informal caregivers to identify the key parameters which need to be addressed to achieve effective at-home lifestyle monitoring systems. Scoring 60%, their strengths come from the inclusion of both medical and end user input, however they similarly lack extensive consideration for how the parameters being set by stakeholders would affect the performance of the monitoring systems themselves. Itoh et al [46] investigated issues of application from a medical perspective but score only 40%. Not only does their investigation consist only of testing an existing technology without the intention of developing said technology further, they also concern their research only with application concerns as far as the caregivers and medical stakeholders are concerned, with no involvement being afforded to the end users and patients. As a result, issues around stigma and even privacy are completely unaddressed.

As can be seen in Table 3, there is a clear lack of involvement in research and application design from the medical community and caregivers. Overall, 13.7% and 64.2% of single-discipline papers involve end user participation versus a mere 14.2% and 0% exhibiting caregiver or medical participation in the development of the research (for application and technical based papers. respectively). While the likelihood of involving both types of stakeholders (caregivers and end users) is more than double when multidisciplinary teams are involved (14.2% vs 37.5%), this is not conclusive due to the relatively small number of papers in each category when divided by the inclusion of different stakeholder types (end user and caregiver). Likewise, there is a definite trend across all research team types that exhibit lower evaluation scores when neither caregivers nor end users participate in the design or evaluation of research. However, discerning the differing impact between the 2 stakeholder types is difficult due to the small positive sample size (with only 2%

separating end user only and caregiver only scores for both research team types, refer to Table 4). The trend in the research, especially in survey-style research seems to be to focus on end users rather than caregivers or other medical stakeholders with only 7 of 60 papers involving medical stakeholders and 22 of 60 involving end user input.

Table . The presence or absence of caregivers and end users in the co-design of research across the review, split by research team type.

Team type	Caregivers (%)	End users (%)	Total (%)
Multidisciplinary[a]	37.5	56.25	*37.5*
Application[a]	38.4	64.2	14.2
Technical[a]	0	13.7	0

[a] Values are not mutually exclusive and some papers have both caregivers and end users involved. As a result, values may not necessarily add up to 100%.

Table . The average evaluation score by research team type when caregivers and end users are included or not in the research.

Team type	Neither (score %)	Caregivers (score %)	End users (score %)	Caregiver and End users (score %)
Multidisciplinary	35.7	70	68	70
Application	27.5	40	42	43
Technical	24.8	0	35	0

The strongest scores using the evaluation framework come from those where interdisciplinary teams are used (average score of 5.43 versus 3.5 and 2.68 for application-based and technical teams respectively), with a distinct trend shown in Table 4 indicating the presence of caregiver and end user involvement in a co-design or PPI capacity leads to a stronger evaluation score.

In general, as shown in Figure 3, different categories of research team tend to neglect certain aspects identified in the evaluation framework. Technical papers struggle with “Stigma” and “Human Response,” and application-based papers, mostly led by medical teams often omit factors such as “Accessibility“ and “Burden to Others” from their consideration. Both tend to struggle particularly with addressing societal stigma, possibly due to the varying levels and types of stigma for at-home monitoring from researcher’s home countries, or potentially due to considerations for more general issues not directly related to their specific research being deemed as outside of the scope of their work.

A common theme across all of these works is the tendency to make assumptions on the necessity of their application in the eyes of end users (Figure 3), as the lack of perceived need is seldom addressed regardless of the make-up of the research team. Technical issues are generally well addressed such as usability and reliability, however broader human issues such as societal stigma and feelings of end users becoming a burden, or the technology itself being a burden to use by caregivers are likewise rarely addressed. Across the entire evaluation phase, there is a consistent trend of greater representation of every attribute when multidisciplinary teams are involved in the research, with specifically technical teams being especially susceptible to a lack of focus on application-based concerns.

Discussion

Principal Findings

The research presented here concludes that a root problem underlying the issues affecting real-life uptake of research applications is the lack of multidisciplinarity in the research. We find that the developed evaluation metric is effective for quantifying the likelihood an application will transition into further development and effective deployment in real-world use. We also identified a series of common themes across the literature, namely the consistent underrepresentation of end users and other stakeholders in research, a distinct lack of multidisciplinarity in technical applications and the consistent underappreciation of societal factors such as stigma and human contact when developing at-home medical technologies.

Using a novel evaluation metric based on the 10 key barriers to uptake identified in the literature, a concrete trend can be observed that that multidisciplinarity between engineers, computer scientists, medical experts, and social scientists, alongside co-design or PPI-based inclusion of caregivers (such as family, friends, or dependants) and end users (health care professionals or patients) is highly beneficial to the effective development of technology that has direct practical application and tangible social benefit. The evaluation framework itself was also justified by the demonstration of a trend of extremely high evaluation scores on research projects that had reached the point of successful practical application, namely HalleyAssist and SPHERE. Using this evaluation metric, we find an increase of between 19.3% - 27.5% for multidisciplinary teams versus single discipline, and an increase of between 7.4% - 19% when research includes the use of caregivers or end users in a PPI or co-design capacity.

Limitations and Future Work

The main limitation of this study is that the results can only demonstrate the lack of multidisciplinarity is a root issue, not necessarily the root issue. For example, many extenuating factors not included in the research itself can explain why effective technology was researched but not developed, such as personal material conditions for the researchers, leading them to be uninterested or unable to develop effective research further. One other limitation is in the methodology of the framework itself. Ways to improve the descriptive power and nuance of the framework could be to introduce more criteria, break the criteria down into further subdisciplines or introduce criteria weighting so that criteria deemed more “vital” or less common are weighted higher when calculating the score. Had more resources been available, a broader review in both phases could have been conducted, using a greater number of papers from a greater number of databases, an especially important factor given the number of different disciplines relevant to this type of research.

In future, a larger study could be conducted on the literature to further verify the effectiveness of the evaluation framework, including the application of the framework to more smart-home technologies in commercial development. The framework itself could also benefit from being imbued with greater complexity for example, additional factors or a bespoke weighting of them to make certain factors more important to the success of a technology in application, and thus scoring a higher score. To our knowledge there is no comparable work in quantitatively assessing the effects of multidisciplinarity on research in at-home health monitoring, so in general, more research is needed to saturate the domain and allow for trends to be more concretely identified. As HalleyAssist and SPHERE are leading examples in the area of in-home monitoring of ageing singles, it would be good to follow up their development and experiences in the future, to see how well the assessment undertaken here was predictive of their successes.

Acknowledgments

This research was funded by the Legal and General Group (research grant to establish the independent Advanced Care Research Centre at University of Edinburgh).

Authors' Contributions

CL and RBF were both involved in the conception of the paper, with CL carrying out the literature reviews and being the primary drafter of the manuscript. RBF was also responsible for supervisory support and redrafting and revising the manuscript. Both authors have given final approval for the manuscript presented for publishing.

Conflicts of Interest

This research was funded by the Legal & General Group (research grant to establish the independent Advanced Care Research Centre at University of Edinburgh). The funder had no role in the conduct of the study, interpretation or the decision to submit for publication. The views expressed are those of the authors and not necessarily those of Legal & General. To the authors knowledge there are no conflicts of interest during the undertaking of this research.

Multimedia Appendix 1
Supplementary materials referred to in the manuscript. There is also a live hyperlink.
[XLSX File, 51 KB - humanfactors_v12i1e59458_app1.xlsx]

Checklist 1
PRISMA checklist.
[PDF File, 182 KB - humanfactors_v12i1e59458_app2.pdf]

References

1. United Nations. World Population Ageing 2019 Highlights: United Nations; 2019. [doi: 10.18356/9df3caed-en]
2. Boström M, Kjellström S, Björklund A. Older persons have ambivalent feelings about the use of monitoring technologies. Info Technol Disab 2013;25(2):117-125. [doi: 10.3233/TAD-130376]
3. Hawley-Hague H, Boulton E, Hall A, Pfeiffer K, Todd C. Older adults’ perceptions of technologies aimed at falls prevention, detection or monitoring: A systematic review. Int J Med Inform 2014 Jun;83(6):416-426. [doi: 10.1016/j.ijmedinf.2014.03.002] [Medline: 24798946]
4. Kaur B, Kumar S, Kaushik BK. Novel wearable optical sensors for vital health monitoring systems-a review. Biosensors (Basel) 2023 Jan 23;13(2):181. [doi: 10.3390/bios13020181] [Medline: 36831947]
5. Lee LN, Kim MJ. A critical review of smart residential environments for older adults with a focus on pleasurable experience. Front Psychol 2019;10:3080. [doi: 10.3389/fpsyg.2019.03080] [Medline: 32038424]

6. Liu L, Stroulia E, Nikolaidis I, Miguel-Cruz A, Rios Rincon A. Smart homes and home health monitoring technologies for older adults: A systematic review. Int J Med Inform 2016 Jul;91:44-59. [doi: 10.1016/j.ijmedinf.2016.04.007] [Medline: 27185508]
7. Majumder S, Mondal T, Deen MJ. Wearable sensors for remote health monitoring. Sensors (Basel) 2017 Jan 12;17(1):130. [doi: 10.3390/s17010130] [Medline: 28085085]
8. Morita PP, Sahu KS, Oetomo A. Health monitoring using smart home technologies: scoping review. JMIR Mhealth Uhealth 2023 Apr 13;11:e37347. [doi: 10.2196/37347] [Medline: 37052984]
9. Shaik T, Tao X, Higgins N, et al. Remote patient monitoring using artificial intelligence: Current state, applications, and challenges. WIREs Data Min Knowl 2023 Mar;13(2). [doi: 10.1002/widm.1485]
10. Wang J, Spicher N, Warnecke JM, Haghi M, Schwartze J, Deserno TM. Unobtrusive health monitoring in private spaces: the smart home. Sensors (Basel) 2021 Jan 28;21(3):864. [doi: 10.3390/s21030864] [Medline: 33525460]
11. Zheng YL, Ding XR, Poon CCY, et al. Unobtrusive sensing and wearable devices for health informatics. IEEE Trans Biomed Eng 2014 May;61(5):1538-1554. [doi: 10.1109/TBME.2014.2309951] [Medline: 24759283]
12. Tao X, Yun Z. Fall prediction based on biomechanics equilibrium using Kinect. Int J Distrib Sens Netw 2017 Apr;13(4):155014771770325. [doi: 10.1177/1550147717703257]
13. Brunzini A, Caragiuli M, Massera C, Mandolini M. Healthy ageing: a decision-support algorithm for the patient-specific assignment of ICT devices and services. Sensors (Basel) 2023 Feb 7;23(4):1836. [doi: 10.3390/s23041836] [Medline: 36850433]
14. Guo Y, Yang J, Liu Y, Chen X, Yang GZ. Detection and assessment of Parkinson's disease based on gait analysis: a survey. Front Aging Neurosci 2022;14:916971. [doi: 10.3389/fnagi.2022.916971] [Medline: 35992585]
15. Kulurkar P, Dixit CK, Bharathi VC, Monikavishnuvarthini A, Dhakne A, Preethi P. AI based elderly fall prediction system using wearable sensors: a smart home-care technology with IOT. Measurement: Sensors 2023 Feb;25:100614. [doi: 10.1016/j.measen.2022.100614]
16. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 1;20(5):e10235. [doi: 10.2196/10235] [Medline: 29716883]
17. Higuma M, Sanjo N, Mitoma H, Yoneyama M, Yokota T. Whole-day gait monitoring in patients with Alzheimer's disease: a relationship between attention and gait cycle. J Alzheimers Dis Rep 2017 Apr 28;1(1):1-8. [doi: 10.3233/ADR-170001] [Medline: 30480224]
18. Colantonio S, Jovanovic M, Zdravevski E, et al. Are active and assisted living applications addressing the main acceptance concerns of their beneficiaries? preliminary insights from a scoping review. Presented at: Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments; Jun 29 to Jul 1, 2022; Corfu Greece. [doi: 10.1145/3529190.3534753]
19. Burrows A, Gooberman-Hill R, Craddock I, Coyle D. Sensors for healthcare: would you want them in your home. arXiv. Preprint posted online on 2014arXiv:1408.2476.
20. Morgan C, Craddock I, Tonkin EL, et al. Protocol for PD SENSORS: Parkinson's disease symptom evaluation in a naturalistic setting producing outcome measuRes using SPHERE technology. an observational feasibility study of multi-modal multi-sensor technology to measure symptoms and activities of daily living in Parkinson's disease. BMJ Open 2020 Nov 30;10(11):e041303. [doi: 10.1136/bmjopen-2020-041303] [Medline: 33257491]
21. Zhu N, Diethe T, Camplani M, et al. Bridging e-Health and the internet of things: the SPHERE project. IEEE Intell Syst 2015;30(4):39-46. [doi: 10.1109/MIS.2015.57]
22. Forkan ARM, Branch P, Jayaraman PP, Ferretto A. HalleyAssist: a personalised internet of things technology to assist the elderly in daily living. 2019 Jan 8 Presented at: Hawaii International Conference on System Sciences. [doi: 10.24251/HICSS.2019.507]
23. Ghorayeb A, Comber R, Gooberman-Hill R. Older adults' perspectives of smart home technology: are we developing the technology that older people want? Int J Hum Comput Stud 2021 Mar;147:102571. [doi: 10.1016/j.ijhcs.2020.102571]
24. McNeill A, Briggs P, Pywell J, Coventry L. Functional privacy concerns of older adults about pervasive health-monitoring systems. Presented at: Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments; 2017; Island of Rhodes Greece. [doi: 10.1145/3056540.3056559]
25. Robinson L, Brittain K, Lindsay S, Jackson D, Olivier P. Keeping In Touch Everyday (KITE) project: developing assistive technologies with people with dementia and their carers to promote independence. Int Psychogeriatr 2009 Jun;21(3):494-502. [doi: 10.1017/S1041610209008448] [Medline: 19193255]
26. Mann WC, Belchior P, Tomita MR, Kemp BJ. Older adults perception and use of PDAs, home automation system, and home health monitoring system. Top Geriatr Rehabil 2007 Jan;23(1):35-46. [doi: 10.1097/00013614-200701000-00006]
27. Austin J, Dodge HH, Riley T, Jacobs PG, Thielke S, Kaye J. A smart-home system to unobtrusively and continuously assess loneliness in older adults. IEEE J Transl Eng Health Med 2016;4:2800311. [doi: 10.1109/JTEHM.2016.2579638] [Medline: 27574577]
28. Moretti G, Marsland S, Basu D, Sen Gupta G, The MUSE Group. Towards a monitoring smart home for the elderly: one experience in retrofitting a sensor network into an existing home. J Ambient Intell Smart Environ 2013;5(6):639-656. [doi: 10.3233/AIS-130233]

29. Mohammed KI, Zaidan AA, Zaidan BB, et al. Real-time remote-health monitoring systems: a review on patients prioritisation for multiple-chronic diseases, taxonomy analysis, concerns and solution procedure. J Med Syst 2019 Jun 11;43(7):223. [doi: 10.1007/s10916-019-1362-x] [Medline: 31187288]
30. Abbas M, Le Bouquin Jeannès R. Acceleration-based gait analysis for frailty assessment in older adults. Pattern Recognit Lett 2022 Sep;161:45-51. [doi: 10.1016/j.patrec.2022.07.006]
31. Lin B, Zhang S, Bao F. Gait recognition with multiple-temporal-scale 3D convolutional neural network. Presented at: Proceedings of the 28th ACM International Conference on Multimedia; 2020; Seattle WA USA. [doi: 10.1145/3394171.3413861]
32. Yu S, Tan D, Tan T. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. Presented at: 18th International Conference on Pattern Recognition (ICPR'06); 2006; Hong Kong, China. [doi: 10.1109/ICPR.2006.67]
33. Sabo A, Mehdizadeh S, Iaboni A, Taati B. Estimating Parkinsonism severity in natural gait videos of older adults with dementia. IEEE J Biomed Health Inform 2022 May;26(5):2288-2298. [doi: 10.1109/JBHI.2022.3144917] [Medline: 35077373]
34. Yan S, Xiong Y, Lin D. Spatial temporal graph convolutional networks for skeleton-Based action recognition. AAAI 2018;32(1). [doi: 10.1609/aaai.v32i1.12328]
35. Goetz CG. Unified Parkinson's disease rating scale (UPDRS) and the movement-disorder society sponsored-unified Parkinson's disease rating scale (MDS-UPDRS). Encyclo Mov Disord 2010:307-309. [doi: 10.1016/B978-0-12-374105-9.00091-5]
36. Yin Z, Jiang Y, Zheng J, Yu H. STJA-GCN: a multi-branch spatial–temporal joint attention graph convolutional network for abnormal gait recognition. Appl Sci (Basel) 2023;13(7):4205. [doi: 10.3390/app13074205]
37. Du Y, Lim Y, Tan Y. A novel human activity recognition and prediction in smart home based on interaction. Sensors (Basel) 2019 Oct 15;19(20):4474. [doi: 10.3390/s19204474] [Medline: 31619005]
38. Islam M, Kabir M, Mridha MF, Alfarhood S, Safran M, Che D. Deep Learning-Based IoT System for Remote Monitoring and Early Detection of Health Issues in Real-Time. Sensors (Basel) 2023;23(11):5204. [doi: 10.3390/s23115204]
39. Megalingam RK, Pocklassery G, Jayakrishnan V, Mourya G, Thulasi AA. Smartphone based continuous monitoring system for home-bound elders and patients. Presented at: 2014 International Conference on Communications and Signal Processing (ICCSP); 2014; Melmaruvathur, India. [doi: 10.1109/ICCSP.2014.6950039]
40. Rachakonda L, Marchand DT. Fall-sense: an enhanced sensor system to predict and detect elderly falls using iomt. Presented at: 2022 IEEE Computer Society Annual Symposium on VLSI (ISVLSI); Jul 4-6, 2022; Nicosia, Cyprus. [doi: 10.1109/ISVLSI54635.2022.00100]
41. Rajaei H. An iot based smart monitoring system detecting patient falls. Presented at: 2022 Annual Modeling and Simulation Conference (ANNSIM); Jul 18-20, 2022; San Diego, CA, USA. [doi: 10.23919/ANNSIM55834.2022.9859346]
42. Mireles C, Sanchez M, Cruz-Ortiz D, Salgado I, Chairez I. Home-care nursing controlled mobile robot with vital signal monitoring. Med Biol Eng Comput 2023 Feb;61(2):399-420. [doi: 10.1007/s11517-022-02712-y]
43. Bjornsdottir K, Ceci C. Home care nursing practice for older persons with heart failure living at home. J Clin Nurs 2023 Jul;32(13-14):4103-4115. [doi: 10.1111/jocn.16575] [Medline: 36369989]
44. Robinson EL, Park G, Lane K, Skubic M, Rantz M. Technology for healthy independent living: creating a tailored in-home sensor system for older adults and family caregivers. J Gerontol Nurs 2020 Jul 1;46(7):35-40. [doi: 10.3928/00989134-20200605-06] [Medline: 32597999]
45. Braspenning AM, Cranen K, Snaphaan L, Wouters EJM. A multiple stakeholder perspective on the drivers and barriers for the implementation of lifestyle monitoring using infrared sensors to record movements for vulnerable older adults living alone at home: a qualitative study. Int J Environ Res Public Health 2022 Jan 5;19(1):570. [doi: 10.3390/ijerph19010570] [Medline: 35010829]
46. Itoh S, Tan HP, Kudo K, Ogata Y. Comparison of the mental burden on nursing care providers with and without mat-type sleep state sensors at a nursing home in Tokyo, Japan: quasi-experimental study. JMIR Aging 2022 Mar 23;5(1):e19641. [doi: 10.2196/19641] [Medline: 35319474]

Abbreviations

CNN: convolutional neural network
PPI: public and participant involvement
ST-GCN: Spatiotemporal Graph Neural Network
STJA-GCN: Spatio-temporal Joint Adjacency Graph Neural Network

Edited by A Kushniruk; submitted 12.04.24; peer-reviewed by M Jordan-Marsh, TA El-Hafeez; revised version received 14.11.24; accepted 24.11.24; published 27.02.25.

Please cite as:
Lochhead C, Fisher RB
On the Necessity of Multidisciplinarity in the Development of at-Home Health Monitoring Platforms for Older Adults: Systematic Review
JMIR Hum Factors 2025;12:e59458
URL: https://humanfactors.jmir.org/2025/1/e59458
doi:10.2196/59458

Challenges to Rehabilitation Services in Sub-Saharan Africa From a User, Health System, and Service Provider Perspective: Scoping Review

Callixte Cyuzuzo[1*], MSc, MPH; Marie Josee Dukuzimana[2*], MSc; Clement Muhire[3*], MSc; Mathew Sheldon Ames[4*], PhD; Emmanuel Ngwakongnwi[2*], PhD

[1]Center for Population Health, Institute of Global Health Equity Research, University of Global Health Equity, Kigali, Rwanda

[2]Institute of Global Health Equity Research, University of Global Health Equity, KG 7 Ave, 5th Floor, PO Box 6955, Kigali, Rwanda

[3]Campus Operations and Community Engagement, University of Global Health Equity, Burera, Rwanda

[4]Educational Development and Quality Center, University of Global Health Equity, Kigali, Rwanda

[*]all authors contributed equally

Corresponding Author:
Marie Josee Dukuzimana, MSc
Institute of Global Health Equity Research, University of Global Health Equity, KG 7 Ave, 5th Floor, PO Box 6955, Kigali, Rwanda

Abstract

Background: Rehabilitation aims to restore and optimize the functioning of impaired systems for people with disabilities. It is an integral part of universal health coverage, and access to it is a human right.

Objective: We aimed to identify the key challenges to rehabilitation services in Sub-Saharan Africa from a user, health system, and service provider perspective.

Methods: This scoping review was conducted in accordance with the 5-stage framework proposed by Arksey and O'Malley. A comprehensive electronic search was run to identify published articles on rehabilitation services in Sub-Saharan Africa. Of the 131 articles retrieved, 83 articles were assessed for eligibility and 15 papers that met the inclusion criteria were considered.

Results: The results show that people with disabilities in Sub-Saharan Africa face multifactorial challenges to access rehabilitation services. Poor access to rehabilitation services is associated with less attention given to rehabilitation by governments, which leads to less funding, negative cultural and social beliefs, fewer rehabilitation centers, poorly equipped rehabilitation units, failure of health systems, lack of training to rehabilitation practitioners, and logistical and financial constraints. This review also reveals that digital rehabilitation reduces costs and improves access to services in hard-to-reach geographical areas. However, digital rehabilitation faces challenges as well, including connectivity issues, inaccessibility to technology, a lack of technical knowledge, a lack of privacy, and ethical concerns.

Conclusions: People with disabilities face multifactorial challenges to access rehabilitation services in Sub-Saharan Africa. It is therefore critical to address these challenges to optimize patients' health outcomes and offer better rehabilitation services.

(JMIR Hum Factors 2025;12:e58841) doi:10.2196/58841

KEYWORDS

challenges; users; health system; service providers; Sub-Saharan Africa; scoping review; rehabilitation service

Introduction

Rehabilitation refers to a set of therapeutic approaches that are structured to address physical, developmental, emotional, and mental challenges and improve patients' health outcomes and quality of life [1]. Rehabilitation is an integral part of universal health coverage and contributes greatly to the achievement of the United Nations' Sustainable Development Goal 3, which aims to ensure healthy lives and promote well-being for people of all ages [2]. Access to rehabilitation for people with disabilities is a human right, as stated in Article 26 of the United Nations Convention on the Rights of Persons with Disabilities [3]. Rehabilitation services can be either delivered through traditional face-to-face methods or, using remote approaches, facilitated by existing technologies [4,5]. Digital rehabilitation has reduced costs and improved access to rehabilitation services in hard-to-reach geographical areas [5].

Demographic changes that lead to chronic health conditions and accidents have gradually contributed to the increased need for rehabilitation services [6]. It is estimated that 2.4 billion people have a health condition that may benefit from rehabilitation services; however, these services are still

inaccessible due to a shortage of rehabilitation practitioners [1]. A World Health Organization (WHO) report disclosed that the ratio of skilled rehabilitation practitioners to patients is 10 to 1,000,000 in low- and middle-income countries. This leads to a persistent scarcity of health-related human resources and reduced access to rehabilitation services in low-resource regions [7].

Sub-Saharan Africa (SSA) has a shortage of health-related human resources, with the region accounting for only 3.5% of the world's health workers, despite it having a high disease burden that includes several kinds of disabilities [1]. In this low-resource region, there is a significant difference in the availability of and access to rehabilitation services, and this impedes individuals with disabilities from achieving their desired health outcomes [8,9]. The WHO reported that in SSA, 50% of people do not get the rehabilitation services they need, and a high proportion of people with disabilities with unmet needs are in low- and middle-income countries [10].

Demand for rehabilitation services in SSA is well established given the prevalence of unmet rehabilitation needs, the rising cases of noncommunicable diseases, and the significant incidence of road traffic accidents that result in disabilities. Additionally, lack of access to rehabilitation services for people with chronic conditions can lead to a need for assistance with activities of daily living and long-term hospital stays [11]. It is important to explore the rehabilitation challenges faced by users, health systems, and service providers. This scoping review aims to enhance our understanding of the complex demands imposed by these factors in SSA and examine the challenges associated with rehabilitation services from those 3 domains.

Methods

Study Design

This review follows the framework of conducting scoping reviews as proposed by Arksey and O'Malley [12]. This framework consists of 5 distinct stages: identifying the research question; identifying relevant studies; selecting studies; charting the data; and collating, summarizing, and reporting the results.

Identifying the Research Question

The research topic and its associated objectives helped to determine the scope of the review, appropriate literature, search strategy, and inclusion and exclusion criteria. To attain the main aim of this review, the scope of rehabilitation services in SSA was explored and key challenges that users, service providers, and health systems face on the availability of and access to rehabilitation services were identified.

Identifying Relevant Studies

Search Strategy

A comprehensive electronic search was conducted to retrieve published data on the availability of and access to rehabilitation services in SSA. Research studies published in reputable journals and databases such as PubMed, National Center for Biotechnology Information, Scopus, PLOS, BioMed Central, Taylor and Francis, Science Direct, Frontiers, Springer Nature, and Web of Science were chosen, as they yielded the most topic-relevant articles. A search strategy using the keywords "rehabilitation services," "availability," "access," and "Sub-Saharan Africa" was used. To increase the number of accessed literatures, "Sub-Saharan Africa" was replaced by the name of the country, and "rehabilitation services" was replaced by specific types of rehabilitation services such as physiotherapy, occupational therapy, speech and language therapy, and prosthetics services. To maximize the literature coverage and to provide additional evidence, reference lists of primary studies and review articles were screened to identify additional relevant literature.

Inclusion and Exclusion Criteria

Inclusion criteria were developed to ensure that all relevant publications focusing on availability of and access to rehabilitation services in SSA were included. Works written in English and published in the above-mentioned reputable journals and databases between 2018 and 2023 were considered. Articles not written in English, published before 2018, and that did not report on rehabilitation services in SSA countries were excluded.

Study Selection

One team member led the search and screened the titles and abstracts of all retrieved papers based on the inclusion and exclusion criteria. Articles retained for review were confirmed by the research team after discussion. Duplicates were removed and the articles that satisfied the inclusion criteria were selected for full-text review.

Charting the Data

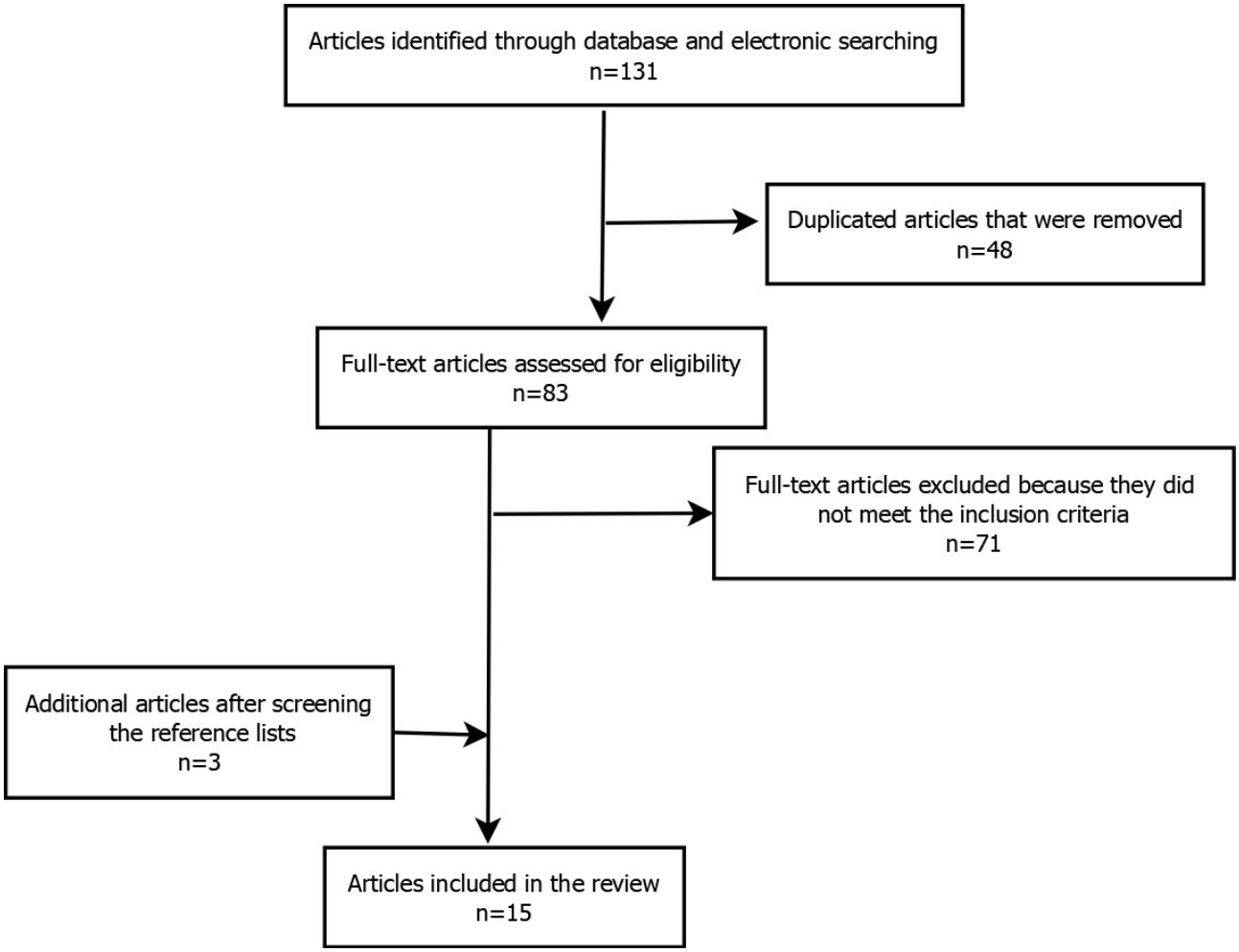
With the help of Microsoft Excel, data on availability of and access to rehabilitation services were extracted from the full texts. Challenges faced by users, service providers, and health systems were recorded. Information such as the title of the paper, names of authors, year of publication, study setting, study design, objectives of the paper, inclusion criteria, and study findings were extracted and recorded in the Excel table. Reference lists of included papers were also screened to optimize the search. The data extraction document is attached as a supplementary file (Multimedia Appendix 1).

Results

Overview

The search yielded 131 research and review articles overall. Forty-eight articles were duplicates and 71 articles did not satisfy the inclusion criteria. The full texts of the remaining 12 articles were screened, and 3 more articles were identified using the reference lists, resulting in a total of 15 articles eligible for analysis. Six of the 15 (40%) articles were original studies while 9 (60%) were scoping or systematic reviews covering other aspects of rehabilitation. Figure 1 illustrates how the articles were selected. All 15 considered articles indicated the existence of rehabilitation challenges; of these, 6 papers discussed the challenges faced by users, 5 discussed those faced by health systems, 3 discussed those faced by service providers, and 1 review article reported challenges for all.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flowchart illustrating the article selection process.



Challenges to Rehabilitation Services in SSA

A synthesis of the literature revealed 2 categories of challenges to rehabilitation services from user, service provider, and health system perspectives (Table 1).

Table . Challenges to rehabilitation services reported in studies involving users, providers, and health systems.

	Category	
	Accessibility	Availability
Users	Overall challenges: • Unsuitable environment and buildings for people with disabilities • Inaccessibility to technology • Stigma, negative attitudes, and discrimination from therapists and other hospital staff • Long queues and waiting time • Financial constraints to purchasing medicines • Mindset that nothing can be done to help them • Lack of psychological support • Nonsupportive family members • Lack of trust in service providers • Pain during the rehabilitation process • Religious beliefs and socially accepted norms and traditions that claim that disability results from witchcraft or divine origins • Language barriers Digital rehabilitation–specific challenges: • Lack of secure platforms and privacy • Lack of technical knowledge or digital literacy • Connectivity issues • Inconsistency in power supply	• Do not know where to find the services • Long distance between users' homes and health facilities • Lack of reliable transport and logistical affordability • Distant health units
Service providers	Overall challenges: • Insufficient time for consultation • Inadequate follow-up • Inappropriate treatment procedures • Lack of patience • Ethical challenges Digital rehabilitation–specific challenges: • Lack of technical knowledge or digital literacy • Connectivity issues	• Lack of communication skills • Lack of professional training and necessary skills • Lack of maintenance for the assistive tools • Failure to schedule online appointments
Health systems	Overall challenges: • Consideration of rehabilitation as a less important health care strategy that is not well integrated into services • Fragmented health services • Lack of community awareness on rehabilitation services • Failure of the health system Digital rehabilitation–specific challenges: • Financial constraints	• Irregular referral to well-equipped health facilities • Lack of medicines and services • Shortage of health-related human resources • Low attention from governments and underfunding • Limited rehabilitation centers and poorly equipped rehabilitation units

Discussion

Principal Findings

This review revealed that people with disabilities in SSA have limited access to rehabilitation services. The limited access can be explained partly by user, service provider, and health system factors [8,13]. Studies included in this review [5,14] found that rehabilitation was given less priority by governments and not recognized as a crucial health care service. Consequently, the sector faced underfunding; experienced a deficiency in health-related human resources, particularly in therapists; encountered a scarcity of rehabilitation centers; and witnessed inadequately equipped rehabilitation units within health facilities. This contributed to the compromised access to high-quality rehabilitation services [5,14,15].

Another study reported that limited access to rehabilitation services in SSA is attributed to the financial inability to afford treatment, lack of insurance covering service expenses, long waiting times at the health facilities, and lack of drugs at health facilities [10]. Other studies have also found that some patients lack psychological support, are unaware of where to find

services, believe that nothing could be done to help them, and are not involved in making decisions on their health [8,16].

This review identified stigma as an impediment to accessing rehabilitation services. A study conducted in Rwanda [17] found that stigma around disability, coupled with negative attitudes and discrimination from therapists and other hospital staff, negatively impacted the uptake of rehabilitation services. Likewise, the role of family in accessing rehabilitation services is well documented. Patients with family members who do not perceive the need for rehabilitation services and who do not trust service providers may be demotivated to seek rehabilitation services [13,17,18]. Additionally, compromised access to health information, irregularities in referral to health facilities that are well equipped, inadequate policies and standards that govern the services, lack of adequate follow-up, insufficient time for consultations, and pain during the rehabilitation process are all hindrances to accessing rehabilitation services [9,15].

This review found that people with disabilities in SSA lack reliable, affordable, and accessible transportation means to reach health facilities. Furthermore, long distances to services and health units, unsuitable environments, and a lack of wheelchair-accessible buildings prevent users from accessing rehabilitation services. A review conducted in Brazil documented patients' determinants (residential location, economic resources, and social characteristics) and characteristics of services (cost, location, and status of the available facilities) as factors that limit access to rehabilitation services [8].

The results published by Bright et al [18] revealed that compromised health systems, lack of professional training and necessary skills for therapists, inappropriate treatment procedures, and lack of maintenance for the assistive tools reduce access to high-quality rehabilitation services. Health care providers who have not received enough training may lack the skills required to provide successful rehabilitation services. This lack of competence implies that misdiagnoses or outdated and inappropriate techniques could be used, causing harm to patients.

The study conducted by Jones et al [5] on telerehabilitation reported how digital rehabilitation services have addressed some logistical challenges; however, they come with additional challenges, including the absence of secure platforms and privacy, inaccessibility to technology, and lack of technical knowledge for both users and therapists [8]. Other challenges identified include a failure to schedule online appointments, connectivity issues, ethical challenges, and failure of the health systems. Patients and health care providers may struggle to use digital tools for rehabilitation, limiting the effectiveness of remote health services. Additionally, the inability to schedule and failure to maintain digital appointments may result in missed sessions, decreased patient participation, and care delays. Health systems face the challenge of poor policies that lead to a lack of financial capacity to afford digital technologies in health service delivery [19]. Digital technologies such as online medical records, wearable devices, and telemedicine can improve patient care and accessibility to health services, particularly in hard-to-reach areas. When financial limitations prevent the use of these technologies, patients may not receive the necessary high-quality care on time.

This review discerned that religious beliefs, cultural norms, and traditions that attribute disabilities to witchcraft or divinity serve as deterrents for individuals with disabilities seeking rehabilitation services at health facilities. A study conducted in Rwanda found that people with mental disorders believe that traditional and faith healers are more effective at treating mental problems than hospital specialists [20]. Such cultural beliefs dissuade people with disabilities from rehabilitation services, casting doubt on their quality and efficacy. Uncertainties regarding the precise origins of their conditions, which are often intertwined with mystical beliefs, add to this dissuasion [21].

Considering the rehabilitation challenges identified in this review, there is a need to bridge the gaps of infrastructure and social and cultural awareness regarding disability and rehabilitation services. This should be done by decentralizing services, providing continuous professional training to therapists, and promoting regular community awareness about rehabilitation services to inform service seekers and break the stigma around it. While this review has enumerated the challenges to rehabilitation services in SSA, it is worth noting that a majority of articles retained for consideration were scoping and systematic reviews that examined other aspects of rehabilitation. User and health system challenges were mostly highlighted in comparison to service provider challenges, which were limited in content.

Conclusion

This review reveals that rehabilitation services in SSA face multifactorial challenges that negatively impact timely access and quality of rehabilitation services for people with disabilities. As a consequence, those who require rehabilitation services experience longer periods of decreased mobility and functioning, inferior quality of life, and lower socioeconomic well-being. Future studies should examine the application of digital technologies to improve rehabilitation services' accessibility, especially in remote settings.

Authors' Contributions

All authors contributed to all phases of data collection and manuscript preparation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data extraction table.
[DOCX File, 19 KB - humanfactors_v12i1e58841_app1.docx]

Checklist 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) checklist.
[PDF File, 110 KB - humanfactors_v12i1e58841_app2.pdf]

References

1. Mapulanga M, Dlungwane T. Mapping evidence of community health workers delivering physical rehabilitation services in sub-Saharan Africa: a scoping review protocol. BMJ Open 2022 Mar 29;12(3):e053736. [doi: 10.1136/bmjopen-2021-053736] [Medline: 35351705]
2. Goal 3: ensure healthy lives and promote well-being for all at all ages. United Nations. URL: https://www.un.org/sustainabledevelopment/health/ [accessed 2025-02-10]
3. Convention on the rights of persons with disabilities and optional protocol. United Nations. URL: https://www.un.org/disabilities/documents/convention/convoptprot-e.pdf [accessed 2025-02-10]
4. Décary S, De Groote W, Arienti C, et al. Scoping review of rehabilitation care models for post COVID-19 condition. Bull World Health Organ 2022 Nov 1;100(11):676-688. [doi: 10.2471/BLT.22.288105] [Medline: 36324552]
5. Jones L, Lee M, Castle CL, Heinze N, Gomes RSM. Scoping review of remote rehabilitation (telerehabilitation) services to support people with vision impairment. BMJ Open 2022 Aug 1;12(8):e059985. [doi: 10.1136/bmjopen-2021-059985] [Medline: 35914903]
6. Jesus TS, Landry MD, Hoenig H. Global need for physical rehabilitation: systematic analysis from the Global Burden of Disease Study 2017. Int J Environ Res Public Health 2019 Mar 19;16(6):980. [doi: 10.3390/ijerph16060980] [Medline: 30893793]
7. Magnusson L, Kebbie I, Jerwanska V. Access to health and rehabilitation services for persons with disabilities in Sierra Leone - focus group discussions with stakeholders. BMC Health Serv Res 2022 Aug 5;22(1):1003. [doi: 10.1186/s12913-022-08366-8] [Medline: 35932077]
8. Clemente KAP, da Silva SV, Vieira GI, et al. Barriers to the access of people with disabilities to health services: a scoping review. Rev Saude Publica 2022;56:64. [doi: 10.11606/s1518-8787.2022056003893] [Medline: 35792776]
9. Naidoo U, Ennion L. Barriers and facilitators to utilisation of rehabilitation services amongst persons with lower-limb amputations in a rural community in South Africa. Prosthet Orthot Int 2019 Feb;43(1):95-103. [doi: 10.1177/0309364618789457] [Medline: 30044179]
10. Tannor AY, Nelson MES, Steere HK, Quao BO, Haig AJ. Building PRM in sub-Saharan Africa. Front Rehabil Sci 2022;3:910841. [doi: 10.3389/fresc.2022.910841] [Medline: 36189021]
11. Treger I, Kosto A, Vadas D, Friedman A, Lutsky L, Kalichman L. Crafting the future of community-based medical rehabilitation: exploring optimal models for non-inpatient rehabilitation services through a narrative review. Int J Environ Res Public Health 2024 Oct 8;21(10):1332. [doi: 10.3390/ijerph21101332] [Medline: 39457305]
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
13. Padmavathi J, Gandhi S, Siva Kumar T. Systematic review on end-users' perception of facilitators and barriers in accessing tele-rehabilitation services. J Psychosoc Rehabil Ment Health 2023 Mar 27;10(3):1-12. [doi: 10.1007/s40737-023-00333-8] [Medline: 37360916]
14. Conradie T, Berner K, Louw Q. Rehabilitation workforce descriptors: a scoping review. BMC Health Serv Res 2022 Sep 17;22(1):1169. [doi: 10.1186/s12913-022-08531-z] [Medline: 36115976]
15. Hussein El Kout NAR, Pilusa S, Masuku KD. A review of the framework and strategy for disability and rehabilitation services in South Africa. Afr J Disabil 2022;11:893. [doi: 10.4102/ajod.v11i0.893] [Medline: 36567920]
16. Louw Q, editor. Collaborative Capacity Development to Complement Stroke Rehabilitation in Africa: AOSIS; 2021. [doi: 10.4102/aosis.2020.BK85]
17. Kumurenzi A, Richardson J, Thabane L, Kagwiza J, Musabyemariya I, Bosch J. Provision and use of physical rehabilitation services for adults with disabilities in Rwanda: a descriptive study. Afr J Disabil 2022;11:1004. [doi: 10.4102/ajod.v11i0.1004] [Medline: 36092478]
18. Bright T, Wallace S, Kuper H. A systematic review of access to rehabilitation for people with disabilities in low- and middle-income countries. Int J Environ Res Public Health 2018 Oct 2;15(10):2165. [doi: 10.3390/ijerph15102165] [Medline: 30279358]
19. Owhor GA, Adeniyi AA, Shomuyiwa D. Overview of digital health in Sub-Saharan Africa: challenges and recommendations. IOSR J Nurs Health Sci 2023;12(1):19-21. [doi: 10.9790/1959-1201041921]
20. Muhorakeye O, Biracyaza E. Exploring barriers to mental health services utilization at Kabutare District Hospital of Rwanda: perspectives from patients. Front Psychol 2021;12:638377. [doi: 10.3389/fpsyg.2021.638377] [Medline: 33828506]

21. Grandpierre V, Milloy V, Sikora L, Fitzpatrick E, Thomas R, Potter B. Barriers and facilitators to cultural competence in rehabilitation services: a scoping review. BMC Health Serv Res 2018 Jan 15;18(1):23. [doi: 10.1186/s12913-017-2811-1] [Medline: 29334968]

Abbreviations

SSA: Sub-Saharan Africa
WHO: World Health Organization

Edited by A Kushniruk, E Borycki; submitted 26.03.24; peer-reviewed by D Franzsen, S Levine; revised version received 04.12.24; accepted 08.12.24; published 28.02.25.

Please cite as:
Cyuzuzo C, Dukuzimana MJ, Muhire C, Sheldon Ames M, Ngwakongnwi E
Challenges to Rehabilitation Services in Sub-Saharan Africa From a User, Health System, and Service Provider Perspective: Scoping Review
JMIR Hum Factors 2025;12:e58841
URL: https://humanfactors.jmir.org/2025/1/e58841
doi:10.2196/58841

AI Chatbots for Psychological Health for Health Professionals: Scoping Review

Gumhee Baek[1,2*], MS, RN; Chiyoung Cha[1,3*], PhD, RN; Jin-Hui Han[1*], MS, RN

[1]College of Nursing, Ewha Womans University, 52 Ewhayeodae-gil, Daehyun-dong, Seodaemun-gu, Seoul, Republic of Korea

[2]Graduate Program in System Health Science and Engineering, Ewha Womans University, Seoul, Republic of Korea

[3]College of Nursing, Ewha Research Institute of Nursing Science, Ewha Womans University, Seoul, Republic of Korea

[*]all authors contributed equally

Corresponding Author:
Jin-Hui Han, MS, RN
College of Nursing, Ewha Womans University, 52 Ewhayeodae-gil, Daehyun-dong, Seodaemun-gu, Seoul, Republic of Korea

Abstract

Background: Health professionals face significant psychological burdens including burnout, anxiety, and depression. These can negatively impact their well-being and patient care. Traditional psychological health interventions often encounter limitations such as a lack of accessibility and privacy. Artificial intelligence (AI) chatbots are being explored as potential solutions to these challenges, offering available and immediate support. Therefore, it is necessary to systematically evaluate the characteristics and effectiveness of AI chatbots designed specifically for health professionals.

Objective: This scoping review aims to evaluate the existing literature on the use of AI chatbots for psychological health support among health professionals.

Methods: Following Arksey and O'Malley's framework, a comprehensive literature search was conducted across eight databases, covering studies published before 2024, including backward and forward citation tracking and manual searching from the included studies. Studies were screened for relevance based on inclusion and exclusion criteria, among 2465 studies retrieved, 10 studies met the criteria for review.

Results: Among the 10 studies, six chatbots were delivered via mobile platforms, and four via web-based platforms, all enabling one-on-one interactions. Natural language processing algorithms were used in six studies and cognitive behavioral therapy techniques were applied to psychological health in four studies. Usability was evaluated in six studies through participant feedback and engagement metrics. Improvements in anxiety, depression, and burnout were observed in four studies, although one reported an increase in depressive symptoms.

Conclusions: AI chatbots show potential tools to support the psychological health of health professionals by offering personalized and accessible interventions. Nonetheless, further research is required to establish standardized protocols and validate the effectiveness of these interventions. Future studies should focus on refining chatbot designs and assessing their impact on diverse health professionals.

(JMIR Hum Factors 2025;12:e67682) doi:10.2196/67682

KEYWORDS

artificial intelligence; AI chatbot; psychological health; health professionals; burnout; scoping review

Introduction

Health professionals are on the front line of patient care, enduring enormous responsibilities and constant pressure. Their work often occurs in an environment in which they must make life-and-death decisions, making them particularly vulnerable to psychological burnout [1]. According to recent studies, more than 50% of health professionals worldwide experience symptoms of burnout, such as fatigue, cynicism, and decreased effectiveness, leading to low job satisfaction [2], high turnover [3], and low-quality patient care [4]. This phenomenon not only causes an individual crisis but also has broader implications for health care systems [2,4]. Since the COVID-19 pandemic, psychological health issues such as anxiety, depression, and burnout among health professionals have become more prevalent because of long working hours, high-stress working environments, and exposure to traumatic events [5]. These issues directly affect health professionals' well-being and work performance and can ultimately negatively impact the quality of health care [6]. Therefore, supporting the psychological health of health professionals and preventing burnout are urgent challenges.

While traditional face-to-face counseling has been used to prevent psychological illness among health professionals, this

method presents various limitations including temporal and spatial constraints, limited availability of mental health professionals, accessibility barriers, and anonymity concerns [7-9]. Health professionals face significant challenges in accessing psychological care owing to demanding work schedules and high work intensity [10]. Additionally, they may feel psychologically burdened when seeking help in situations in which anonymity is not guaranteed [8]. To address these challenges, artificial intelligence (AI) chatbots have emerged as a potential solution, offering a new way to deliver psychological health interventions that can compensate for the limitations of traditional methods [7,8]. AI chatbots have the potential to be accessible anytime and anywhere while maintaining user anonymity, which may provide an alternative approach to psychological health support [7].

Recently, rapid advances in AI and natural language processing (NLP) have opened new possibilities for psychological health interventions, with AI chatbots gaining particular attention in the field of psychological health. Studies have demonstrated the potential of AI chatbots—AI-driven conversational agents—in providing mental health support [11] while emphasizing the importance of their safe and explainable implementation [12]. AI chatbots analyze users' conversation histories and data to provide advice and support tailored to their specific needs. This can help health professionals better manage psychological health issues, such as burnout, and enable timely interventions tailored to individual situations [13]. Previous research has shown that AI chatbots can improve coping skills through natural conversations with users [7], enhance the user experience, facilitate effective intervention delivery [13,14], and provide interventions tailored to specific needs [13,14]. This conversational approach enhances the effectiveness of interventions by fostering positive user engagement, as people are more likely to empathize with and respond to tools that feel engaging and relatable [6,7,14].

Despite the potential of AI chatbots for enhancing psychological health, there remains a lack of research on their effectiveness for health professionals. Much of the existing research has focused on using AI chatbots in patient care, whereas relatively little research exists on the use of AI chatbots to support health professionals. This research gap highlights the need to systematically investigate the characteristics and effectiveness of AI chatbots that target health professionals. Therefore, this study aimed to fill this gap and analyze the characteristics and effectiveness of AI chatbot interventions targeting health professionals to contribute to future research and practice.

Methods

Study Design

We conducted a scoping review to identify the available evidence regarding AI chatbot interventions for health professionals' psychological health. A scoping review maps the relevant literature in the field and is useful for identifying gaps in knowledge [15]. This review was conducted in accordance with the scoping review framework of Arksey and O'Malley. The study adopted the following five steps: (1) identifying the research question; (2) identifying relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results. The five stages provided a structured approach to assist in understanding the current evidence by examining the literature. There was no rating of quality performed on the included studies, as scoping studies do not require a quality evaluation [16]. Therefore, a scoping review is suitable for providing a broader perspective on the current research and mapping the relevant literature on AI chatbot interventions for health professionals' psychological health.

Identifying the Research Question and Relevant Studies

The research question was, "What is the status of the current body of knowledge regarding AI chatbots for the psychological health of health professionals?" According to Arksey and O'Malley, the search strategy involves identifying relevant studies, including published and unpublished works, with an emphasis on being as comprehensive as possible [15]. A literature search was performed between May and August 2024 from the following eight databases: seven global search engines in English (Cochrane Library, CINAHL, Embase, ProQuest Dissertations & Theses Global, PsyclNFO, PubMed, and Web of Science) and one Korean search engine in Korean (Research Information Sharing Service). Various combinations of search terms (Textbox 1) were used to either broaden or narrow the search, depending on the results in a specific database. The reference lists of the retrieved studies were searched manually. Finally, gray literature, relevant letters, and editorials were searched.

Textbox 1. Query box

medical team OR medical staff* OR healthcare worker* OR healthcare provider* OR health personnel* OR health professional* OR doctor* OR surgeon* OR physician* OR nurse* OR midwife* OR Licensed Practical Nurse* OR licensed vocational nurse* OR physician assistant OR nurse practitioner OR clinical nurse specialist* OR nurse clinician* OR registered nurse* OR resident* OR therapist* OR pharmacist* OR nurse assistant* OR nurse aid* AND burnout OR burn-out* OR depress* OR depression OR stress* OR exhaust* OR depersonalization OR personal accomplishment OR anxiety* AND chatbot* OR artificial intelligence OR conversational agent* OR dialogue system* OR machine learning OR natural language process*

Study Selection

Overview

Studies published before May 2024 were selected for this review. The studies retrieved from the literature search were reviewed by three researchers using the following inclusion and exclusion criteria.

Inclusion Criteria

(1) The purposes of the studies included the development or implementation of AI chatbot interventions. (2) Participants of the studies included licensed health professionals such as

physicians or surgeons, dentists, medical doctors, registered nurses, licensed practice or vocational nurses, pharmacists, and allied health professionals such as respiratory and physical therapists. Studies that included nonhealth professionals in addition to health professionals as part of their participants were included if the main purpose of the AI chatbot was to alleviate the psychological distress of health professionals. (3) The intention of the AI chatbots was to improve the psychological health of health professionals, focusing on outcomes such as stress, burnout, depression, and anxiety. (4) The studies were written in English or Korean.

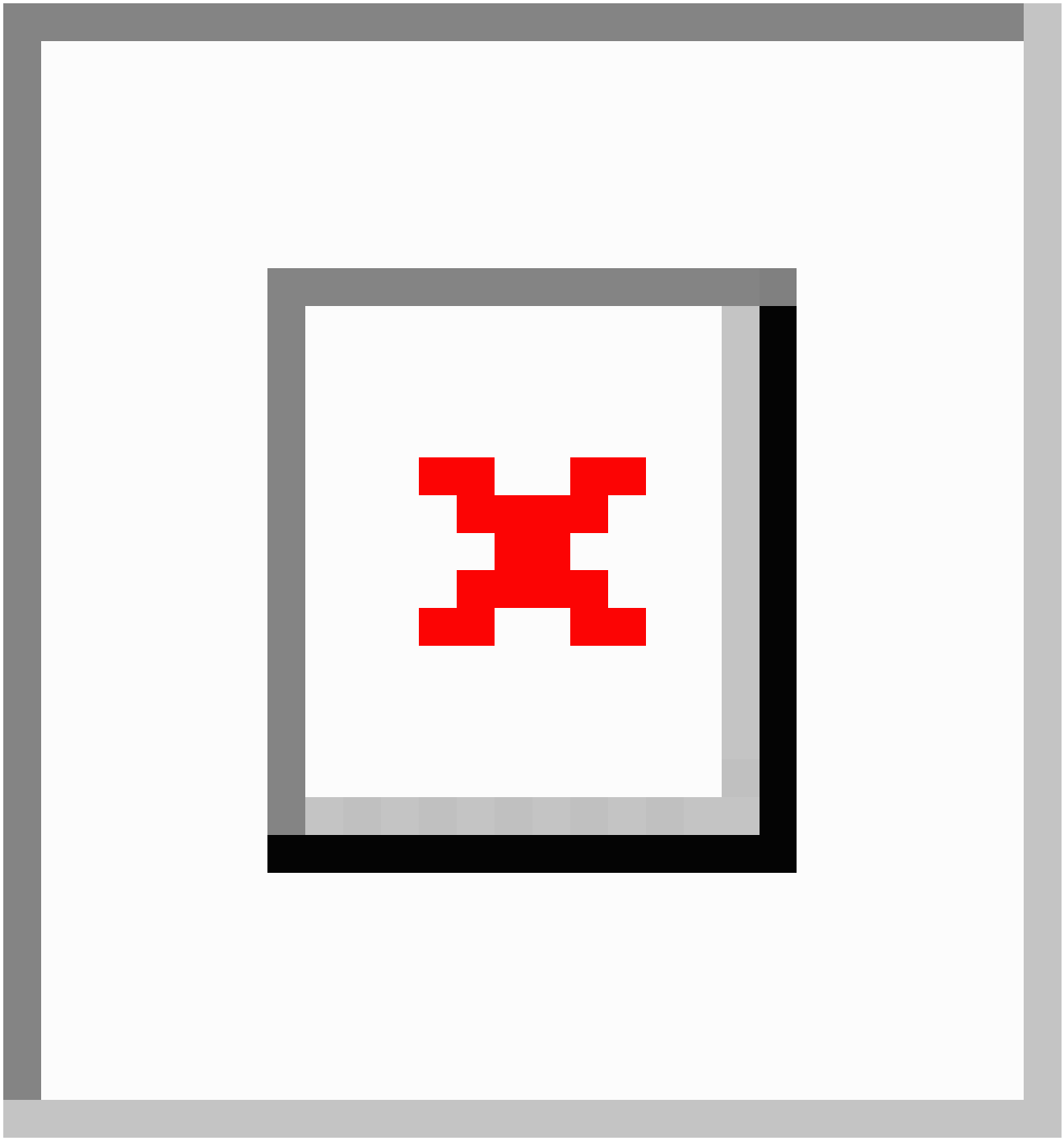
Exclusion Criteria

Studies in which we could not retrieve the full text were excluded.

A total of 2465 studies were retrieved from the initial database search (Figure 1). After removing duplicates, 1766 studies remained. The titles and abstracts were reviewed, and 1750 studies were excluded, leaving 16 studies for a full-text review. Six studies completely met the inclusion criteria after full-text review. Then, by reviewing the reference lists and manually searching, five additional studies were obtained, of which four studies met the criteria for the scoping review. Finally, 10 studies were included in this scoping review.

Throughout the review process, two researchers (GB and JH) took the lead in data extraction, and all researchers reviewed the content. Any uncertainties and disagreements were resolved through weekly meetings. The literature review software Covidence (Veritas Health Innovation Ltd), a study-managing tool for collaborating researchers [17], was used to manage the retrieved studies and to screen for redundant studies.

Figure 1. Flowchart. AI: artificial intelligence.



Charting the Data

The data are presented in tables, summarized, and synthesized qualitatively according to the purpose and research questions of the review.

Collating, Summarizing, and Reporting the Results

The final stage of a scoping review involves collating, summarizing, and reporting the results [15]. Data extraction was performed by two researchers (GB and JH) and checked for accuracy by all the researchers. Uncertainties and disagreements were resolved through discussions. The data were extracted using a common data extraction format.

Results

General Characteristics

Table 1 presents the characteristics of the analyzed studies. Ten studies published between 2018 and 2024 were retrieved. The studies were published worldwide: in Canada [7,18], Japan [19], Malawi [20], Netherlands [21], Singapore [13,22], Spain [23], and the United States [24,25]. Four studies examined diverse health professionals, including doctors, nurses, nurse assistants, clinical officers, pharmacists, lab technicians, and other allied health professionals [20,21,23,26]. Two studies specifically targeted doctors and nursing staff [13,27] such as nurse practitioners, registered nurses, licensed practical nurses, and nursing assistants, while one study exclusively covered chatbots

for nurses [19]. Three studies identified their participants as health professionals, without specifying the specific types [7,18,22]. Of the ten studies, eight studies focused solely on health professionals [13,19-23,26,27]. Two studies included both health professionals and their families [7,18].

The study designs were diverse. Of the 10 studies, three were development studies [18,23,27], one was a design verification study [26], one was a pilot study [22], one was a randomized controlled trial [20], one was a mixed-method feasibility study [21], one was an intervention study [19], one was a cross-sectional study protocol study [7], and one was a qualitative study [13]. Three studies evaluated the development and feasibility of AI chatbots for psychological health support [23,26,27].

Table . Study characteristics.

Author (year)/ country	Study design	Participants	Source of participants	Sample size (M:F[a] ratio)	Age (years), mean (SD)	Publication type
Anmella et al [27](2023)/Spain	Development study	Primary care and health care professionals; medical doctors (GPs[b] and specialists from diverse disciplines), nurses, and nurse assistants	• GP referrals to PCMHSP[c] • Hospital Clínic de Barcelona health professionals	• Simulation study: 17 (24:76) • Feasibility and effectiveness study: 34 (24:76)	• Simulation study: 36.5 (9.7) • Feasibility and effectiveness study: 35.3 (10.12)	Journal study
Chaudhry and Islam [26](2021)/United States	Design verification study	Health professionals in their workplaces during the COVID-19 pandemic; RNs[d], NPs[e], LPN[f], CNAs[g], PAs[h], RT[i], PTs[j], students or intern nurses	Local medical services departments and health professionals' mailing lists	22 (9.1:90.9)	34.98 (8.55)	Journal study
Chang et al [22] (2024)/Singapore	Pilot exercise (service evaluation study)	Health professionals (types unspecified)	National tertiary health care cluster	527 (N/A)[k]	N/A	Journal study
Jackson-Triche et al [23] (2023)/United States	Development and usability study	Health professionals including faculty, staff, and trainees across various disciplines within the health care system	UCSF[l]	Over 26,000 (N/A)	N/A	Journal study
Kleinau et al [20] (2024)/Malawi	Randomized controlled trial	Health professionals (doctors, nurses, clinical officers, medical assistants, pharmacists, lab technicians, physiotherapy technicians)	Public and private health care facilities within Blantyre and Lilongwe districts	1584 (N/A) Completed: 511 (E=296, C=215)	N/A	Journal study
Kroon [21](2018)/Netherlands	Mixed method feasibility study (one group pre-posttest design)	Health professionals (radiation expert, social worker, nurse in training, training officer, pedagogue, health professional disability care HRD[m] adviser	Email using personal and iThrive connections.	12 (8.3:91.7) (interview participant=9)	Half of the participants were aged 18 to 25 years	Master's thesis
Matsumoto et al [19](2021)/Japan	Intervention study	Nurses	Two hospitals in Tokyo	70 (N/A)	N/A	Journal study
Noble et al [7](2022)/Canada	Cross-sectional study protocol	Health professionals and their families	The Canadian provinces of Alberta and Nova Scotia	2200 (N/A)	N/A	Journal study
Yoon et al [13] (2021)/Singapore	Qualitative study	Frontline health professionals (doctors and nurses)	Singapore general hospital and community care facilities (large scale institutional isolation units)	42 (24:76)	29.6 (3.9)	Journal study
Zamani [18](2022)/Canada	Development study	Health professionals and their families	The Canadian provinces of Alberta and Nova Scotia	N/A	N/A	Master's thesis

[a]M:F: male:female.
[b]GP: general practitioner.
[c]PCMHSP: Primary Care Mental Health Support Program.
[d]RN: registered nurse.
[e]NP: nurse practitioner.
[f]LPN: licensed practical nurse.
[g]CNA: certified nurse assistant.
[h]PA: physician assistant.
[i]RT: respiratory therapist.
[j]PT: physical therapist.
[k]N/A: not available.
[l]UCSF: University of California, San Francisco
[m]HRD: human resource development.

AI Chatbot Characteristics

Table 2 shows the characteristics of the AI chatbots developed or implemented in the retrieved studies. The types of AI chatbot algorithms were diverse; NLP-based AI was the most commonly used algorithm, appearing in six studies [7,13,18,19,21,23]; rule-based AI, an algorithmic method that does not learn from data or create rules on its own, was used in two studies [20,22]; and learning-based AI that learns from data was adopted in two studies [26,27]. AI chatbots were delivered through mobile apps in six studies [13,19,20,22,26,27], and web-based platforms via web links in four studies [7,18,21,23]. The interaction type in all studies was 1:1 interaction: Among them, one study included 1:1 interaction alongside group chat functionality [26], and another study included 1:1 interaction with optional human coach support [22]. In seven studies, the chatbot actively initiated the conversation [7,13,18-20,23,27], while in three studies, the conversation was passively initiated [21,22,26].

Participant completion of the AI chatbot interventions was reported in four studies [20,21,23,27]; among these, three studies provided the percentage of participants who completed the intervention: 14.3% [21], 67.7% [20], and 73% [27] while one study reported the employee access rate of the intervention as 10.88% [27]. Of the 10 studies, six evaluated the AI chatbot's usability [13,20,22,23,26,27]. Four studies relied on participant feedback: 54.6% rated it as useful [26], the average satisfaction score was 4.07 out of five [22], more than 80% reported it as helpful [23], and 91% of the users found the chatbot easy to use [20]. One study used user engagement indicators to identify high subjective usability (acceptability, usability, and satisfaction) but low objective engagement (completion, adherence, compliance, and engagement) [27]. Another study assessed chatbot performance metrics, reporting an intent detection accuracy of 99.1% and an entity extraction accuracy of 95.4% for the Mira Chatbot [18]. Chatbot usability was not evaluated in the remaining four studies [7,13,19,21].

Table . AI[a] chatbot characteristics.

Author (year)/ Country	Chatbot name	Purposes of chatbot	AI chatbot type	AI algorithm	Mode of delivery	Interaction type	Initiates conversation	Completion related	Chatbot usability evaluation
Anmella et al [27] (2023)/Spain	Vickybot	To screen anxiety, depression, and burnout with interventions	Digital decision support platform (PRESTO[b])	• Machine-learning severity prediction model	Mobile app	1:1 interaction	Active	27% completed second self-assessment	• High subjective UEI[c] (acceptability, usability, and satisfaction). • Low objective UEI (completion, adherence, compliance, and engagement)
Chaudhry and Islam [26] (2021)/United States	iChatBot (AIO) PeerChat (AI X)	To provide stress management support during COVID-19	Intelligent chatbot	• iChatBot: machine learning model • PeerChat: immediate communication with expert peers	Mobile app	iChatBot: 1:1 PeerChat: group chat	Passive (on-demand queries)	Not reported	• 54.6% participants rated the app as either useful or very useful • 59.1% participants were willing to use it
Chang et al [22] (2024)/Singapore	Wysa	To evaluate Wysa for health professionals' support	Conversational chatbot	• Rule-based AI model	Mobile app	1:1 interaction with optional human coach support	Passive	Not reported	• High engagement rate with positive feedback (91-92%) • Mean feedback score of 4.07 (SD 0.95) out of 5

Author (year)/ Country	Chatbot name	Purposes of chatbot	AI chatbot type	AI algorithm	Mode of delivery	Interaction type	Initiates conversation	Completion related	Chatbot usability evaluation
Jackson-Triche et al [23]/United States	UCSF[d] Cope	To provide behavioral health support via chatbot	An algorithm-based, automated, web-based AI conversational tool	• Natural language model	Web-based interface	1:1 interaction	Active	10.88% of employees accessed the chatbot	• >80%of attendees reported the experience as helpful
Kleinau et al [20] (2024)/Malawi	Viki (Vitalk's virtual mental health care assistant)	To assess chatbot impact on mental health in Malawi	Automated conversational agent	• Rule-based AI model	Mobile app	1:1 interaction	Active	Of 1584 participants, 511 completed	• 91% found the app easy to use; 87% found the content relevant; 92% benefited from the app, and 83% felt more resilient • Common barriers: lack of time (45%), app duration (30%), and usability issues (13%)
Kroon [21] (2018)/Netherlands	iThrive	To help health professionals manage inner critics and enhance self-compassion	Preprogrammed conversations and interaction	• Natural language model	Web-based interface	1:1 interaction	Passive (users initiate interaction with the chatbot)	12 of 14 participants included in the final analysis	• N/A[e]
Matsumoto et al [19] (2021)/Japan	CB[f] app (a chatbot app)	To support mental health self-care for busy professionals, especially nurses	Digital decision support platform	• Natural language model	Mobile app, VR[g]	1:1 interaction	Active	Not reported	• N/A
Noble et al [7](2022)/Canada	MIRA[h]		Hybrid NLP[i] and decision tree AI chatbot	• Natural language model	Web-based platform	1:1 interaction	Active	Not reported	• N/A

Author (year)/ Country	Chatbot name	Purposes of chatbot	AI chatbot type	AI algorithm	Mode of delivery	Interaction type	Initiates conversation	Completion related	Chatbot usability evaluation
		To provide tailored mental health information and services for health professionals and their families							
Yoon et al [13](2021)/Singapore	mHealth[j] app	To support frontline health professionals's psychosocial well-being during COVID-19	N/A	• Natural language model	Mobile app	1:1 interaction	Active	Not reported	• N/A
Zamani [18] (2022)/Canada	MIRA	To offer strategic mental health resources for health professionals and their families	Hybrid NLP and decision tree AI chatbot	• Natural language model	Web-based platform	1:1 interaction	Active	Not reported	• Intent detection: 99.1% accuracy in identifying user needs • Entity extraction: 95.4% accuracy in recognizing key terms for resource recommendations.

[a]AI: artificial intelligence.

[b]PRESTO: primary care digital support tool in mental health.

[c]UEI: user engagement indicator.

[d]UCSF cope: University of California, San Francisco faculty, staff, and trainee coping and resiliency program.

[e]Not available.

[f]CB: cognitive behavioral.

[g]VR: virtual reality

[h]MIRA: mental health intelligent information resource assistant.

[i]NLP: natural language processing.

[j]mHealth: mobile health.

AI Chatbot Intervention Characteristics and Outcomes

Table 3 lists the contents and outcomes of the AI chatbot interventions. The primary component is the screening function for mental health conditions such as anxiety, depression, and stress, which was included in seven studies [19-23,26,27]. Four studies included cognitive behavioral therapy techniques [13,20,22,27]. Additionally, four studies implemented stress management techniques, such as breathing exercises and meditation [13,19,20,22]. Three studies incorporated problem-solving strategies [7,18,22] and psychoeducation [19,22,27]. Two studies used behavioral activation [20,22], cognitive restructuring [21,22], and gratitude training [20,22]. One study featured storytelling [21]. In addition, there was a

function to support portal sites connected to community resources in three studies [7,18,23].

Table . AI[a] chatbot intervention characteristics and outcomes.

Author (year)/ country	Chatbot contents	Additional features	Intervention doses	Main outcomes
Anmella et al [27](2023)/Spain	• Screening and monitoring for self-assessment tools • Personalized psychological modules: CBT[b], mindfulness, DBT[c], and ACT[d] • Psychoeducation and coping strategies for stress management • Chatbot-guided navigation	• Emergency alert for suicidal thoughts • Weekly objective reminders • Audio recording option for reflections after self-assessments	• 1 month (self-assessments conducted at baseline and biweekly)	• Anxiety (t_8=1.000; P=.34) and Depressive (t_8=0.40; P=.70) symptoms were not significant differences • Work-related burnout was moderately reduced (z=–2.07, P=.04, r=0.32)
Chaudhry and Islam [26] (2021)/United States	• iChatBot—provides real-time answers (self-monitoring and stress-tracking information related to patient care and workplace stress management) • PeerChat—enables peer-to-peer support	• AssignedTasks: records stress levels and task completion times • MyNotifications: Sends reminders for health activities like sleep, exercise, and meals • SmartMonitor: connects with devices to track vitals (eg, blood pressure, oxygen, pulse, and temperature) • InfectionCheck: Manual COVID-19 symptom checker via chatbot	• N/A[e]	• Usefulness and willingness: 54.6% found the app useful; 59.1% willing to use it. • Six feature feedback: MyNotifications: useful reminders; PeerChat: positive for peer support; AssignedTasks: good for tracking, but time-consuming; InfectionCheck: useful, potential redundancy; StressMonitor: very helpful for coping; iChatBot: concerns about emergency responsiveness
Chang et al [22](2024)/Singapore	• Mental health assessments: anxiety—GAD-7[f], depression—PHQ[g]-2 • Intervention modules: mindfulness, sleep meditation, guided visualization, thought recording, behavioral activation, psychoeducation, breathing exercises, cognitive restructuring, acceptance, grounding, social support, problem-solving, habit building, gratitude training	• Option to connect with a human coach • Weekly objective reminders and notifications • Gratitude challenge to boost engagement • Audio recording for self-reflection	• Varied among users (on average, users completed 10.9 sessions over 3.80 weeks)	• 495 completed at least one full session; 422 completed two or more sessions • The interventions most used were for sleep and anxiety, with a strong repeat-use rate • 46.2% reported symptoms of anxiety • 15.2% reported symptoms of depression
Jackson-Triche et al [23] (2023)/United States	• Mental health assessment screening: suicide or self-harm thoughts, substance use concerns, emotional distress, social withdrawal and work-related struggles, resource navigation • Web-based self-care resources • Behavioral health services: emergency support	• Telehealth assessments for crisis evaluation • Self-management website offering curated resources • In-person navigator for personalized support • Support groups with regular peer meetings	• Varied based on user interaction	• UCSF[h] cope reached all units • 10.88% (3785/34,790) of employees accessed the technology • 39.7% (708/1783) of employees with psychological distress requested in-person services • Positive feedback on all program elements. • Over 80% of attendees found the experience helpful

Author (year)/ country	Chatbot contents	Additional features	Intervention doses	Main outcomes
Kleinau et al [20](2024)/Malawi	• Mental health assessments: anxiety (GAD-7), depression (PHQ-9), burnout (OLBI[i]), loneliness (UCLA Loneliness Scale)[j], resilience (RS-14)[k] • Careline programs (thematic mental health conversation tracks) • Intervention: CBT techniques, positive psychology strategies, behavioral activation, gratitude exercises, breathing, relaxation, and meditation exercises,	• Mood tracking with emojis for human-like interaction • Gamification for engagement • Emergency support with local psychologist contact info for high-risk users	• 8 weeks -First 4 weeks: daily interactions • Second 4 weeks: interactions every other day	• GAD-7 showed a decrease (–0.44), but the result was not statistically significant. • OLBI showed a decrease (–0.58), with borderline significance. • PHQ-9 showed a significant decrease. • UCLA loneliness scale: no significant • RS-14 showed a significant increase.
Kroon [21](2018)/Netherlands	• Awareness of inner critics: storytelling, encouraging reflection • Identification of inner critics: assessment tool • Cognitive restructuring techniques • Self-compassion development	• Humor: adds humor for engaging interactions • Tokens: rewards for task completion • Compliments: encourages users with motivating feedback	• 2 weeks	• ProQOL[l] and SCS[m] no significant. • Quantitative results: no significant changes in self-compassion or compassion fatigue. • Qualitative results: Positive feedback on the iThrive intervention, with increased awareness of inner critics and changes in thinking patterns. • Chatbot interactions reportedly reduced stress and improved self-compassion.
Matsumoto et al [19](2021)/Japan	• Digital-SAT[n] method: self-guided stress management, techniques to convert stress-related physical discomfort into positive sensations, image-based therapeutic exercises • Mental health assessments: self-esteem, anxiety, emotional support, depression, problem-solving abilities • Stress management: guided stress relief sessions, emotional stabilization exercises, visualization techniques	• Automated daily messages with reminders and psychoeducational content • Gamification elements such as progress tracking • Immersive VR[o] experiences	• 4 weeks	• Self-esteem and family support significantly improved. • Self-repression showed a trend of improvement but was not significant. • SDS scores and counseling needs increased in the VR + CB[p] group. • Maximal blood pressure decreased significantly in the VR group and showed a decreasing trend in the VR + CB group. • Minimal blood pressure significantly decreased in the VR group.
Noble et al [7](2022)/Canada	• Mental health education and information: problem-solving • Personalized resource navigation • Self-help strategies and coping skills		• N/A	• The findings will be reported in follow-up research

Author (year)/ country	Chatbot contents	Additional features	Intervention doses	Main outcomes
		• Emergency contacts for urgent mental health support • Hybrid AI approach: free-text input and predefined options • Crisis support referrals for suicidal thoughts		
Yoon et al [13](2021)/Singapore	• Personalized goal setting: tailored goals for lifestyle improvements (eg, sleep, exercise), feedback on progress to encourage behavioral changes, reminders to support adherence to personal goals • Educational resources: mindfulness exercises and short wellness studies, stress management techniques	• In-app counseling and peer support • Reminders and notifications • Gamification elements: point-based rewards • Engaging features to promote continued app use • Mood tracking and progress monitoring	• N/A	• Mood-tracking reminders were helpful, but workers struggled with self-awareness. • Goal-setting and resources were valued; frequent notifications were distracting. • A built-in counselor chat was preferred for accessibility and privacy. • Gamification was not well-received. • Peer support was needed, but app-based interactions raised concerns. • Motivation, usability, and rewards were crucial for app use.
Zamani [18](2022)/Canada	• Information on mental health topics, including substance use and coping strategies: problem-solving • Delivers personalized recommendations • Facilitates access to local mental health services and programs through a resource portal	• Emergency contact option for users experiencing distress • Stores conversation logs • Feedback system	• N/A	• Intent detection: 99.1% accuracy in identifying user needs. • Entity extraction: 95.4% accuracy in recognizing key terms for resource recommendations.

[a]AI: artificial intelligence.

[b]CBT: cognitive behavior therapy.

[c]DBT: dialectical behavioral therapy.

[d]ACT: acceptance and commitment therapy.

[e]N/A: not available.

[f]GAD-7: general anxiety disorder-7.

[g]PHQ: patient health questionnaire.

[h]UCSF cope: University of California, San Francisco faculty, staff, and trainee coping and resiliency program

[i]OLBI: Oldenburg burnout inventory.

[j]UCLA: University of California, Los Angeles loneliness scale.

[k]RS-14: 14-item resilience scale.

[l]ProQOL: Professional Quality of Life Scale.

[m]SCS: Self-Compassion Scale

[n]SAT: structured association technique.

[o]VR: virtual reality.

[p]VR-CB: virtual reality-chatbot.

Across the studies, a variety of additional features were incorporated. Seven studies included AI chatbot functionalities for crisis management and support [7,13,18,20,22,27]. Specific features encompassed emergency contact information for crises [7,18,20,27], peer support groups and in-app forums for social support [13], expert support through in-app counseling, and options for connecting with human coaches [13,22,23]. Six studies incorporated reminders to encourage chatbot use [13,18,19,22,26,27]. Gamification elements, such as point-based rewards to encourage participation, were featured in four studies [13,19-21]. Peer support functionalities were available in three studies [13,23,26]. Two studies included mood-tracking features for health and lifestyle management [13,20], while another two studies provided an audio recording option [22,27]. Additionally, one study integrated devices for tracking vital signs [26] and provided an immersive virtual reality experience [19]. A total of 4 out of 10 studies explicitly stated the duration of chatbot use or intervention dose ([19]: 4 wk; [20]: 8 wk; [21]: 2 wk; [23]: 1 mo) while it was not specified in six studies. The dose varied among users in two studies [22,23].

The research in four studies evaluated psychological health outcomes [19-21,27], whereas, in six studies, the focus was primarily on the development and description of user perceptions of the AI chatbot without providing specific outcome measures [7,13,18,22,23,26]. The most commonly measured psychological health outcomes were anxiety and depression, which were examined in three studies [19,20,27]. Burnout was investigated in two studies [20,27]. Compassion fatigue [21], loneliness [20], resilience [20], and self-compassion [21] were also evaluated. Anxiety was measured in three studies: one reported a significant reduction of scores [20] while the other two reported no significant differences in anxiety scores [19,27]. Burnout was assessed in two studies, and each reported a significant decrease in its scores [20,27]. Depressive symptoms were measured in two studies: one reported a significant reduction of scores [20] while the other did not [27]. Other psychological variables were measured in one study each; of those, the variables that showed significant change in scores were health counseling needs [19] and resilience [20].

Discussion

Principal Results

The studies retrieved for this scoping review were published during the past seven years, between 2018 and 2024, and reflect the rapidly evolving trend of AI chatbot technology for supporting psychological health. The global distribution of studies across countries indicates a widespread interest in using AI for psychological health support and health professionals' psychological health.

Our review revealed that chatbots were used by a broad spectrum of health professionals, highlighting their diverse applications. This diversity underscores the potential for universally designed chatbots to effectively address psychological health needs across different professional settings. However, prior research has shown that nurses, who have direct and continuous patient contact, experience greater mental health distress compared to other health professionals [24]. Therefore, future studies should consider incorporating role-specific options into chatbot designs to create tailored interventions for diverse professional groups.

The diverse research methods used in the retrieved studies signify the nascent stage of research on AI chatbots for psychological health support, emphasizing technology refinement and user experience comprehension. Such a variety of research approaches is crucial for validating the efficacy and safety of AI chatbot technology, enabling iterative improvements and a deeper understanding of how health professionals interact with AI chatbots [25]. However, it is noteworthy that only 4 out of 10 studies measured psychological health outcomes. This limitation underscores the significant gap between technological developments and clinical validation. As the field matures, it becomes imperative to quantitatively assess the impact of AI chatbots on psychological health indicators. Further studies measuring outcome variables to demonstrate the clinical effectiveness of AI chatbots are needed.

The studies reviewed here demonstrate the variety of AI algorithms used in chatbot design, including NLP-based-, rule-, and learning-based approaches. NLP-based AI chatbots which provide personalized responses by analyzing user inputs [28] are the most commonly used AI algorithms. Evidence has shown that an NLP-based AI chatbot, which simulates natural conversations, is successful at delivering tailored interventions [29]. Rule-based AI algorithms, adopted in 2 studies out of the 10 retrieved studies, are effective in providing fixed and predetermined responses; however, they lack the adaptability required to provide more dynamic and engaging psychological interventions [30]. Previous research has shown that NLP technology, which interprets psychological health cues from user-generated texts, significantly improves the effectiveness of interventions [31,32], a finding consistent with our review [7,13,18,19,21,23]. Moreover, the development of deep learning techniques has led to the widespread use of general-purpose pretrained language models in various NLP applications [33]. This suggests that NLP-based algorithms are particularly suitable for delivering personalized psychological support in real-time interactions. Machine learning and deep learning algorithms show promise in enhancing AI chatbots' effectiveness by continuously learning through user interactions. These algorithms can provide accurate and personalized interventions over time [34]. Future studies should investigate how these advanced algorithms can be used to offer more precise psychological support tailored to individual user experiences.

User engagement is a critical factor in the success of mental health interventions; however, several studies did not report intervention completion rates [20,21,23,27]. These issues might become more pronounced in real-world settings, where factors such as interface complexity, the chatbot's ability to provide meaningful and personalized feedback, and technical issues impact the overall user satisfaction and perceived effectiveness of the intervention [35]. This underscores the need for further exploration of how chatbot interactions can be designed to maintain user engagement, possibly through improved user experiences and personalization techniques [36,37]. From a different perspective, unlike traditional interventions, which are typically designed to be completed from start to finish, chatbot

interactions often do not follow this linear structure. Users may terminate interactions once their immediate concerns are addressed, reflecting the unique flexibility of chatbots [38]. For instance, if a chatbot provides critical information for crisis management and the user responds affirmatively, the user is likely to exit the chat after their need is met. This immediate resolution highlights the efficiency of chatbot interventions in addressing user concerns promptly, even if the interaction appears incomplete. Recognizing this distinction is essential for accurately interpreting dropout rates and evaluating chatbot effectiveness.

The AI chatbots reviewed here comprised various therapeutic approaches. The incorporation of established psychological treatment techniques such as cognitive behavioral therapy, mindfulness, and dialectical behavior therapy into AI chatbots represents a novel approach to delivering psychological health interventions. This approach is considered novel because it uses AI to deliver evidence-based psychological techniques 24/7, serve multiple users simultaneously, tailor interventions based on individual responses, and combine various therapeutic modalities on a single platform.

This multidisciplinary therapeutic approach has the potential to address health professionals' diverse psychological needs [31]. AI chatbot interventions are characterized by a digital and responsive design, which allows for integrating practical tools such as stress management techniques, mood tracking, and personalized goal setting [39]. These features align with the immediate and personalized support often required by health professionals. However, while personalization is a key advantage of AI chatbots, a need also exists for some level of standardization to ensure the interventions' quality and effectiveness [40]. This standardization should focus on establishing evidence-based core components and assessment methods while still allowing for personalized delivery [31]. Future research should focus on identifying the most effective components of these chatbots and tailoring them to the specific psychological health needs of health professionals while maintaining a balance between standardization and personalization.

In this study, a temporary increase in depression scores was observed in one study [20], whereas previous research has reported AI chatbots' positive effects in reducing anxiety, depression, and work-related burnout among health professionals [41]. These contrasting findings may reflect the challenges faced by such professionals, who frequently experience high levels of stress and emotional demands [42]. This suggests the complexity of psychological interventions and highlights the importance of careful monitoring to determine the appropriate duration and intensity of chatbot use.

Some studies in our review demonstrated the potential of AI chatbots to act as first-line responders in identifying and managing mental health crises. Prior research has also shown the potential benefits of AI chatbots not only in improving mental and emotional well-being but also in promoting behavioral changes [43]. While these features may only benefit a subset of users, their impact on preventing escalation to severe mental health conditions or emergencies is invaluable. Future research should prioritize the integration and evaluation of these features to enhance the safety and efficacy of chatbot interventions.

Although AI chatbots address the unique time constraints and stress among health professionals, balancing automated responses with genuine therapeutic interactions remains a significant challenge. Personalization technologies that provide tailored features to AI chatbots have emerged as key factors in their effectiveness, with studies reporting increased user engagement and satisfaction through customized responses and immediate support [29]. Consequently, the development of tailored intervention strategies that consider the individual needs and preferences of health professionals is essential to maximize AI chatbots' effectiveness. These strategies should integrate ongoing user feedback and dynamically adapt interventions to maintain long-term user engagement and effectiveness. Future research should focus on larger sample sizes, extended follow-up periods, and diverse health care professional groups of health professionals. Such comprehensive studies will play a crucial role in developing evidence-based guidelines for implementing AI chatbots in psychological health and in advancing the sophistication of personalization algorithms.

Limitations

This scoping review has several limitations. First, the limited number of studies adopting diverse study designs and using various AI algorithms complicated making direct comparisons among the retrieved studies. Second, the findings' generalizability is restricted because many studies were published in specific regions, particularly high-income countries. Additionally, the short duration of many interventions and limited follow-up periods make it difficult to evaluate the long-term effectiveness and sustainability of AI chatbots. Future research should address these limitations by conducting more rigorous trials, standardizing the outcome measures, and exploring the long-term applications of AI chatbots across diverse populations and settings.

Conclusions

This scoping review explored the current state of AI chatbots aimed at supporting the psychological health of health professionals. Although the reviewed studies demonstrated AI chatbots' potential to reduce stress, anxiety, depression, and burnout, research in this area remains in its early stages. The diversity of study design, AI algorithms, therapeutic approaches, and outcome measures highlights this field's innovative but fragmented nature.

Despite promising results, particularly with NLP-based chatbots, a significant need exists for more rigorous and standardized studies to fully evaluate their clinical efficacy. Challenges such as usability issues and limited generalizability of the findings must be addressed to enhance the long-term application and effectiveness of AI chatbots in real-world settings. Future research should focus on refining chatbot designs, expanding the research to diverse populations, and conducting long-term studies to clarify the role of AI chatbots in supporting the psychological health of health professionals. Thus, AI chatbots could become a valuable solution for addressing the

psychological health challenges faced by health professionals worldwide.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00351194).

Authors' Contributions

CC conceptualized the study, acquired funding, and oversaw project administration. GB and JH were responsible for data curation and methodology development. CC, GB, and JH conducted the formal analysis. GB created the visualizations. CC, GB, and JH collaboratively wrote the original draft, and all authors contributed to reviewing and editing the manuscript. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

Checklist 1
PRISMA-ScR checklist. The checklist has been completed in accordance with the journal's guidelines and includes all the items required for a systematic scoping review.
[PDF File, 98 KB - humanfactors_v12i1e67682_app1.pdf]

References

1. Moss M, Good VS, Gozal D, Kleinpell R, Sessler CN. An official Critical Care Societies Collaborative statement: burnout syndrome in critical care health care professionals: a call for action. Am J Crit Care 2016 Jul;25(4):368-376. [doi: 10.4037/ajcc2016133] [Medline: 27369038]
2. Aronsson G, Theorell T, Grape T, et al. A systematic review including meta-analysis of work environment and burnout symptoms. BMC Public Health 2017 Mar 16;17(1):264. [doi: 10.1186/s12889-017-4153-7] [Medline: 28302088]
3. Maslach C, Schaufeli WB, Leiter MP. Job burnout. Annu Rev Psychol 2001;52(1):397-422. [doi: 10.1146/annurev.psych.52.1.397] [Medline: 11148311]
4. Bakker AB, Costa PL. Chronic job burnout and daily functioning: a theoretical analysis. Burn Res 2014 Dec;1(3):112-119. [doi: 10.1016/j.burn.2014.04.003]
5. Smiley RA, Lauer P, Bienemy C, et al. The 2017 National Nursing Workforce survey. J Nurs Regul 2018 Oct;9(3):S1-S88. [doi: 10.1016/S2155-8256(18)30131-5]
6. Gamble A. Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective. AJIM 2020 Jun 2;72(4):509-523. [doi: 10.1108/AJIM-11-2019-0316]
7. Noble JM, Zamani A, Gharaat M, et al. Developing, implementing, and evaluating an artificial intelligence-guided mental health resource navigation chatbot for health care workers and their families during and following the COVID-19 pandemic: protocol for a cross-sectional study. JMIR Res Protoc 2022 Jul 25;11(7):e33717. [doi: 10.2196/33717] [Medline: 35877158]
8. Fortuna KL, Naslund JA, LaCroix JM, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 3;7(4):e16460. [doi: 10.2196/16460] [Medline: 32243256]
9. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935. [doi: 10.2196/15935] [Medline: 32130167]
10. Grover S, Dua D, Shouan A, Nehra R, Avasthi A. Perceived stress and barriers to seeking help from mental health professionals among trainee doctors at a tertiary care centre in North India. Asian J Psychiatr 2019 Jan;39:143-149. [doi: 10.1016/j.ajp.2018.12.020] [Medline: 30639973]
11. Chaudhry BM, Debi HR. User perceptions and experiences of an AI-driven conversational agent for mental health support. Mhealth 2024;10:22. [doi: 10.21037/mhealth-23-55] [Medline: 39114462]
12. Sarkar S, Gaur M, Chen LK, Garg M, Srivastava B. A review of the explainability and safety of conversational agents for mental health to identify avenues for improvement. Front Artif Intell 2023;6:1229805. [doi: 10.3389/frai.2023.1229805] [Medline: 37899961]
13. Yoon S, Goh H, Nadarajan GD, et al. Perceptions of mobile health apps and features to support psychosocial well-being among frontline health care workers involved in the COVID-19 pandemic response: qualitative study. J Med Internet Res 2021 May 31;23(5):e26282. [doi: 10.2196/26282] [Medline: 33979296]
14. Omarov B, Zhumanov Z, Gumar A, Kuntunova L. Artificial intelligence enabled mobile chatbot psychologist using AIML and cognitive behavioral therapy. IJACSA 2023;14(6):137-146. [doi: 10.14569/IJACSA.2023.0140616]

15. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
16. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(69):69. [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
17. Babineau J. Product review: Covidence (systematic review software). J Can Health Libr Assoc 2014;35(2):68. [doi: 10.5596/c14-016]
18. Zamani A. Developing a mental health virtual assistance (chatbot) for healthcare workers and their families. : University of Alberta; 2022. [doi: 10.7939/r3-dwgt-jz32]
19. Matsumoto A, Kamita T, Tawaratsumida Y, et al. Combined use of virtual reality and a chatbot reduces emotional stress more than using them separately. J Univers Comput Sci 2021;27(12):1371-1389. [doi: 10.3897/jucs.77237]
20. Kleinau E, Lamba T, Jaskiewicz W, et al. Effectiveness of a chatbot in improving the mental wellbeing of health workers in Malawi during the COVID-19 pandemic: a randomized, controlled trial. PLoS ONE 2024;19(5):e0303370. [doi: 10.1371/journal.pone.0303370] [Medline: 38805444]
21. Kroon YHE. The appreciation of the ithrive chatbot and its impact on self-compassion and compassion fatigue in healthcare workers: a mixed method feasibility study. : University of Twente; 2018 URL: https://purl.utwente.nl/essays/75151
22. Chang CL, Sinha C, Roy M, Wong JCM. AI-led mental health support (Wysa) for health care workers during COVID-19: service evaluation. JMIR Form Res 2024 Apr 19;8:e51858. [doi: 10.2196/51858] [Medline: 38640476]
23. Jackson-Triche M, Vetal D, Turner EM, Dahiya P, Mangurian C. Meeting the behavioral health needs of health care workers during COVID-19 by leveraging chatbot technology: development and usability study. J Med Internet Res 2023 Jun 8;25:e40635. [doi: 10.2196/40635] [Medline: 37146178]
24. Lai J, Ma S, Wang Y, et al. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. JAMA Netw Open 2020 Mar 2;3(3):e203976. [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
25. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 1;25(9):1248-1258. [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
26. Chaudhry B, Islam A. Design validation of a workplace stress management mobile app for healthcare workers during COVID-19 and beyond. : Springer; 2021 Presented at: Proceedings of EAI MobiQuitous 2021. [doi: 10.1007/978-3-030-94822-1_17]
27. Anmella G, Sanabra M, Primé-Tous M, et al. Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout in primary care and health care professionals: development, feasibility, and potential effectiveness studies. J Med Internet Res 2023 Apr 3;25:e43293. [doi: 10.2196/43293] [Medline: 36719325]
28. Lim SM, Shiau CWC, Cheng LJ, Lau Y. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. Behav Ther 2022 Mar;53(2):334-347. [doi: 10.1016/j.beth.2021.09.007] [Medline: 35227408]
29. Inkster B, Sarda S, Subramanian V. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR Mhealth Uhealth 2018 Nov 23;6(11):e12106. [doi: 10.2196/12106] [Medline: 30470676]
30. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO)—concept and usability test. IEEE Trans Emerg Topics Comput 2020;9(3):1170-1182. [doi: 10.1109/TETC.2020.2974478]
31. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464. [doi: 10.1177/0706743719828977] [Medline: 30897957]
32. Thunström AO, Carlsen HK, Ali L, Larson T, Hellström A, Steingrimsson S. Usability comparison among healthy participants of an anthropomorphic digital human and a text-based chatbot as a responder to questions on mental health: randomized controlled trial. JMIR Hum Factors 2024 Apr 29;11:e54581. [doi: 10.2196/54581] [Medline: 38683664]
33. Ouali Y, Hudelot C, Tami M. An overview of deep semi-supervised learning. arXiv. Preprint posted online on Jun 9, 2020 URL: https://arxiv.org/abs/2006.05278
34. Chakraborty C, Bhattacharya M, Pal S, Lee SS. From machine learning to deep learning: advances of the recent data-driven paradigm shift in medicine and healthcare. Curr Res Biotechnol 2024;7:100164. [doi: 10.1016/j.crbiot.2023.100164]
35. Agnihotri A, Bhattacharya S. Chatbots' effectiveness in service recovery. Int J Inf Manage 2024 Jun;76:102679. [doi: 10.1016/j.ijinfomgt.2023.102679]
36. Xu Y, Niu N, Zhao Z. Dissecting the mixed effects of human-customer service chatbot interaction on customer satisfaction: an explanation from temporal and conversational cues. J Retail Consum Serv 2023 Sep;74:103417. [doi: 10.1016/j.jretconser.2023.103417]
37. Abd-Alrazaq AA, Alajlani M, Ali N, Denecke K, Bewick BM, Househ M. Perceptions and opinions of patients about mental health chatbots: scoping review. J Med Internet Res 2021 Jan 13;23(1):e17828. [doi: 10.2196/17828] [Medline: 33439133]
38. Ashfaq M, Yun J, Yu S, Loureiro SMC. I, Chatbot: modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. Telematics and Informatics 2020 Nov;54:101473. [doi: 10.1016/j.tele.2020.101473]
39. Balcombe L. AI chatbots in digital mental health. Informatics 2023;10(4):82. [doi: 10.3390/informatics10040082]

40. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. J Med Internet Res 2023 Feb 24;25:e40789. [doi: 10.2196/40789] [Medline: 36826990]
41. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]
42. Tamminga SJ, Emal LM, Boschman JS, et al. Individual-level interventions for reducing occupational stress in healthcare workers. Cochrane Database Syst Rev 2023 May 12;5(5):CD002892. [doi: 10.1002/14651858.CD002892.pub6] [Medline: 37169364]
43. Bragazzi NL, Crapanzano A, Converti M, Zerbetto R, Khamisy-Farah R. The impact of generative conversational artificial intelligence on the lesbian, gay, bisexual, transgender, and queer community: scoping review. J Med Internet Res 2023 Dec 6;25:e52091. [doi: 10.2196/52091] [Medline: 37864350]

Abbreviations

AI: artificial intelligence
NLP: natural language processing

Edited by A Kushniruk; submitted 17.10.24; peer-reviewed by C Pankhurst, EM Turner; revised version received 02.02.25; accepted 14.02.25; published 19.03.25.

Please cite as:
Baek G, Cha C, Han JH
AI Chatbots for Psychological Health for Health Professionals: Scoping Review
JMIR Hum Factors 2025;12:e67682
URL: https://humanfactors.jmir.org/2025/1/e67682
doi:10.2196/67682

An AI-Based Clinical Decision Support System for Antibiotic Therapy in Sepsis (KINBIOTICS): Use Case Analysis

Juliane Andrea Düvel[1], MSc; David Lampe[2], MSc; Maren Kirchner[2], BSc; Svenja Elkenkamp[2], MEd, MSc; Philipp Cimiano[3], Prof Dr; Christoph Düsing[3], MSc; Hannah Marchi[4], MSc; Sophie Schmiegel[4], MSc; Christiane Fuchs[4], Prof Dr; Simon Claßen[5], MD; Kirsten-Laura Meier[5], MD; Rainer Borgstedt[6], Dr; Sebastian Rehberg[6], Prof Dr; Wolfgang Greiner[2], Prof Dr

[1]Centre for electronic Public Health Research (CePHR), School of Public Health, Bielefeld University, P.O. Box 10 01 31, Bielefeld, Germany

[2]AG 5 - Health Economics and Health Care Management, School of Public Health, Bielefeld University, Bielefeld, Germany

[3]AG Semantic Computing, Technical Faculty, Bielefeld University, Bielefeld, Germany

[4]Data Science Group, Faculty of Business Administration and Economics, Bielefeld University, Bielefeld, Germany

[5]Department of Anaesthesiology, Surgical Intensive Care, Emergency Medicine, and Pain Therapy, Hospital Bielefeld, University Hospital Bielefeld, Bielefeld, Germany

[6]Department of Anaesthesiology, Intensive Care, Emergency Medicine, Transfusion Medicine and Pain Therapy, Campus Bielefeld-Bethel, University Hospital Bielefeld, Bielefeld, Germany

Corresponding Author:
Juliane Andrea Düvel, MSc
Centre for electronic Public Health Research (CePHR), School of Public Health, Bielefeld University, P.O. Box 10 01 31, Bielefeld, Germany

Abstract

Background: Antimicrobial resistances pose significant challenges in health care systems. Clinical decision support systems (CDSSs) represent a potential strategy for promoting a more targeted and guideline-based use of antibiotics. The integration of artificial intelligence (AI) into these systems has the potential to support physicians in selecting the most effective drug therapy for a given patient.

Objective: This study aimed to analyze the feasibility of an AI-based CDSS pilot version for antibiotic therapy in sepsis patients and identify facilitating and inhibiting conditions for its implementation in intensive care medicine.

Methods: The evaluation was conducted in 2 steps, using a qualitative methodology. Initially, expert interviews were conducted, in which intensive care physicians were asked to assess the AI-based recommendations for antibiotic therapy in terms of plausibility, layout, and design. Subsequently, focus group interviews were conducted to examine the technology acceptance of the AI-based CDSS. The interviews were anonymized and evaluated using content analysis.

Results: In terms of the feasibility, barriers included variability in previous antibiotic administration practices, which affected the predictive ability of AI recommendations, and the increased effort required to justify deviations from these recommendations. Physicians' confidence in accepting or rejecting recommendations depended on their level of professional experience. The ability to re-evaluate CDSS recommendations and an intuitive, user-friendly system design were identified as factors that enhanced acceptance and usability. Overall, barriers included low levels of digitization in clinical practice, limited availability of cross-sectoral data, and negative previous experiences with CDSSs. Conversely, facilitators to CDSS implementation were potential time savings, physicians' openness to adopting new technologies, and positive previous experiences.

Conclusions: Early integration of users is beneficial for both the identification of relevant context factors and the further development of an effective CDSS. Overall, the potential of AI-based CDSSs is offset by inhibiting contextual conditions that impede its acceptance and implementation. The advancement of AI-based CDSSs and the mitigation of these inhibiting conditions are crucial for the realization of its full potential.

(JMIR Hum Factors 2025;12:e66699) doi:10.2196/66699

KEYWORDS

CDSS; use case analysis; technology acceptance; sepsis; infection; infectious disease; antimicrobial resistance; clinical decision support system; decision-making; clinical support; machine learning; ML; artificial intelligence; AI; algorithm; model; analytics; predictive models; deep learning; early warning; early detection

Introduction

Sepsis infections caused by pathogens with antimicrobial resistance (AMR) represent a significant global challenge in health care [1,2]. In 2017, there were 48.9 million new cases of sepsis and 11 million deaths related to sepsis, accounting for 19.7% of all global deaths [3]. In Germany, sepsis incidence increased by an average of 5.7% per year, from 280 cases in 2010 to 370 cases in 2015 per 100,000 individuals [4]. A recent meta-analysis indicated that the 30-day mortality rate for sepsis in Germany was estimated to be 26.5%, which is consistent with the observed rates in North America and Europe [5]. Furthermore, over 1.27 million deaths per year are attributed to AMR worldwide. In 2019, there were 9650 deaths attributable to AMR (mortality rate of 5 per 100,000) and 45,700 deaths associated with AMR (mortality rate of 22 per 100,000) in Germany [6]. Recent data from Germany show heterogeneous trends in AMR proportions of infected patients underscoring the urgent need for enhanced infection prevention measures to limit AMR spread [7]. Inappropriate antibiotic prescribing represents a significant contributing factor [8]. In particular, the prolonged or improper use of nonspecific, broad-spectrum antibiotics is highly problematic, as these antibiotics provide symptomatic relief but also facilitate the development of resistance in other bacteria [9,10].

Clinical decision support systems (CDSSs) provide a solution for promoting targeted and guideline-based antibiotic prescribing [11-13]. CDSSs adopt a variety of forms and are integrated into routine practice. Despite their diverse nature, they share a common objective: to assist medical professionals in identifying the most appropriate form of therapy for each patient, based on existing data and established guidelines, through the use of programmed algorithms or artificial intelligence (AI) [14,15]. In the context of CDSSs, previous studies have demonstrated a decrease in antibiotic prescription rates, more rapid initiation of appropriate antibiotic therapy for patients, and improved clinical outcomes, such as reduced mortality, increased antibiotic-free days, and fewer medical complications [16-19]. However, CDSSs do not always improve clinical practice [20]. For example, there is still insufficient evidence to conclusively show positive effects on therapy duration, dosage, or adherence to clinical guidelines [11,16,21].

Despite the growing body of research on AI-based CDSSs in intensive care units (ICUs) [22,23], there is a notable gap in the implementation of AI tools in routine care in general [24] and the availability of AI-based CDSSs for sepsis in the German health care system. The KINBIOTICS (translated as “AI-based decision support for antibiotic therapy”) project, funded by the German Federal Ministry of Health, aimed to train AI-based algorithms and improve the prediction of suitable antibiotics for sepsis infections on the basis of a comprehensive data set. In addition to the development of an AI-based CDSS, a cross-sectoral resistance observatory and a new rapid test for sequencing the antibiotic genome were developed [25].

Engaging clinicians in the design and development of CDSSs is often suggested as a strategy to enhance the alignment between the system and the needs of its users [26-28]. Interviews and expert groups are approaches used for preimplementation clinician involvement [29]. Insights into clinicians’ views provide valuable information on the barriers and facilitators that affect their willingness to adopt and use CDSSs in practice [30]. By examining clinicians’ perceptions, the study seeks to inform targeted strategies to improve the CDSS’s design, usability, and relevance, thereby promoting more effective and widespread adoption in clinical practice.

This study aimed to perform a use case analysis of a pilot version of the AI-based CDSS within an ICU setting. The objective was to analyze the clinical decision-making processes, and adopt a more comprehensive perspective on the factors that facilitate or hinder the future implementation of such a system.

Methods

Reporting is based on the Standards for Reporting Qualitative Research (SRQR) [31]. The SRQR checklist is provided in Multimedia Appendix 1.

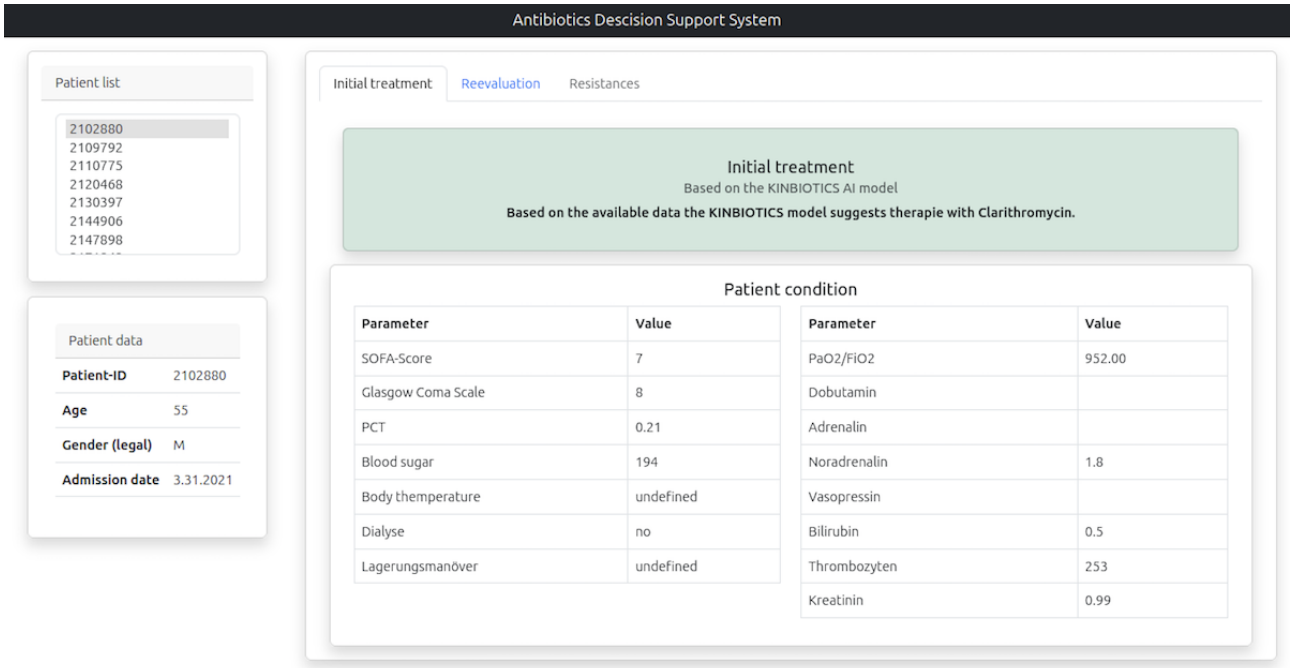
Study Design and Setting

A 2-stage base model (random forest at both stages) was developed for the initial therapy using patient data, laboratory data, and clinical data. Subsequently, the model was optimized using a variety of parameters, including the number of decision trees and tree depth (the model specifications and results will be published separately once the AI-based CDSS model has been finalized). The developed model was trained and tested on data from 1 of the 3 participating clinics and then evaluated for robustness using data from a second clinic. Although the performance of the underlying AI model was insufficient in terms of accuracy, specificity and sensitivity at the time of the evaluation, it was crucial at this time to gain insights into the decision-making mechanisms of the physicians. This was undertaken in order to scrutinize the variables included in the initial model and, where necessary, to supplement or adjust them. Accordingly, the preliminary prototype was not modified during the interview phase. However, subsequent to the interviews, the model underwent alterations and is still under development.

The pilot version of the AI-based CDSS was evaluated in a 2-stage semistructured qualitative process [32]. With regard to feasibility, physicians were each shown 5 exemplary sepsis patient cases in a desktop version of the CDSS. Cases were selected randomly from the evaluation dataset and were identical for each interviewee. In addition to basic patient information, this contained relevant vital parameters, and the treatment recommendation determined by the AI-based CDSS (Figure 1). This antibiotic therapy recommendation was presented to the physicians both for the admission situation to the ICU (initial treatment; the start of antibiotic therapy) and for a possible therapy correction at the time the microbiological findings were available (re-evaluation; usually 48-72 hours after start of antibiotic therapy). The physicians were informed during the individual interviews that the CDSS is still in a pilot status and that the recommendations are likely not yet reliable. The physicians were asked to assess and evaluate the intensive care situation at both time points as well as the AI-based model

recommendation for antibiotic therapy. In addition to the plausibility of the content of the therapy recommendations, the design and layout of the pilot version were evaluated.

Figure 1. Screenshot of the pilot version of the artificial intelligence–based clinical decision support system.



Following the individual interviews, physicians from each clinic were interviewed again as part of a focus group interview. The theoretical focus here was on technology acceptance of AI-based CDSSs and the identification of challenges and conductive conditions with regard to future implementation.

Expert Selection

Expert interviews were conducted with intensive care physicians from the 3 clinical centers involved in the project at the University Medical Center East Westphalia-Lippe. The interviews were conducted on a voluntary basis and the potential participants were selected by the respective project partners in the 3 clinical centers.

Data Collection

Between October and November 2023, the face-to-face expert interviews and focus groups were conducted. The interview guideline (Textbox 1) for the focus group interviews was developed on the basis of the extended Unified Theory of Acceptance and Use of Technology (UTAUT 2) [33,34]. The UTAUT has recently being used to explain technology adoption among health care practitioners' intention to use AI-based CDSSs [35].

Textbox 1. Interview guideline for the focus groups (based on the Unified Theory of Acceptance and Use of Technology 2).

- Opening question
- What were the objectives of today's meeting?
- Questions about the tool
- Please describe your impressions of the KINBIOTICS user interface.
- What information may have been lacking in your medical decision-making process that led to the prescription of antibiotics?
- Please indicate which aspects you consider to be beneficial for a potential application.
- What additional functions would be beneficial to include in the KINBIOTICS user interface to ensure effective utilization in routine clinical practice?
- Questions regarding the use of artificial intelligence (AI) in general
- To what extent do digital tools contribute to the processes undertaken in your day-to-day work?
- Please indicate which digital applications you currently use in your professional activities.
- Please indicate which digital health applications you use in your private life.
- Please indicate your opinion on the utilization of AI, and provide your expectations regarding the implementation of AI in healthcare, both in general and on your ward.
- Please describe your previous experience of contact with AI and clinical decision support systems.
- What is the prevailing attitude in your clinical setting, on the ward and among your professional colleagues with regard to the utilization of artificial intelligence in everyday medical procedures?
- In the hypothetical scenario of utilizing such a tool like KINBIOTICS in tomorrow's clinical decision-making process, what potential challenges might be identified?
- Questions about future utilization of a clinical decision support system
- Please describe the benefits you perceive in the utilization of AI and clinical decision support systems in your everyday professional practice.
- In order to apply such a system in medical decision-making, what prerequisites must be met?
- Please indicate whether you identify any additional aspects that should be addressed for use in everyday clinical practice.

The interviews were recorded by audio and subsequently transcribed. After complete transcription, the audio recordings were irrevocably deleted in accordance with the data protection policy.

Data Analysis

The data analysis of the transcribed interview material was conducted deductively and inductively based on categories according to Kuckartz (2018) [36]. The procedure for the individual interviews primarily followed evaluative content analysis, while the focus groups were analyzed in terms of content structure. To gain as much knowledge as possible, the transcripts of the individual interviews were also analyzed with regard to possible content-structuring findings. The transcripts were analyzed anonymously using the qualitative analysis software MAXQDA (version 2022, VERBI) [37]. The interview analyses and category assignment were carried out independently by 2 researchers (JAD and MK). In the case of discrepancies, a third researcher (DL) was consulted. Categories were adapted or reformulated by the interdisciplinary research team (JAD, MK, DL, SE, and WG).

Ethical Considerations

Informed consent was obtained before the interviews. A positive ethics vote (November 22, 2021) from the ethics committee of the Medical Association of Westphalia-Lippe and the Westphalian Wilhelms University of Münster was received for the entire project (No. 2021-699-f-S). The interview transcripts were anonymized. The interviewees did not receive any compensation; participation was on a voluntary basis.

Results

Sample

A total of 19 individual interviews and 3 focus group interviews were conducted. The distribution of respondents with regard to gender, age, and professional experience was relatively balanced, with 10 (53%) of the respondents identifying as male. In each case, 7 (37%) respondents were aged 30-40 years, and 40-50 years. In total, 5 (26%) respondents had less than 5 years or 11 - 20 years of professional experience, 4 (21%) had between 5 and 10 years, and 3 (16%) respondents indicated that they had more than 20 years of professional medical experience. The duration of the expert interviews ranged from 17 to 43

minutes, while the average duration of the focus group interviews was 1 hour.

Feasibility of the AI-Based Clinical Decision Support System

The approval or rejection of the AI-based antibiotic recommendations was heterogeneous and primarily dependent on 2 factors: first, the plausibility of the AI-based individual therapy (change) recommendations (refer to quotes 1 and 2). It should be noted that not every detailed recommendation was already plausible or medically advisable due to the pilot status of the CDSS respective example cases.

> ***Quote 1:****The system now suggests piperacillin/tazobactam. Of course, this is a very far-fetched suggestion for primary antibiotics when it comes to designing a calculated antibiotic therapy. For me, this would not be the first drug of choice for patients who arrive at the hospital as primary patients, who have been treatment-naïve so far and have not yet received antibiotic therapy [...].* [Expert 5, Clinical Center 3]

> ***Quote 2:****So, [piperacillin/tazobactam]. Well, difficult [...] with a completely normal PCT [Procalcitonin], I would find it difficult [...] to start directly with such a broad antibiotic.* [Expert 8, Clinical Center 2]

> ***Quote 3:****[...] So, [...] such a recommendation, I'm very sure, will be malpractice in the next three to five years [...]. This reflects the reality, but this is the worst thing that can happen, to treat a Staphylococcus aureus bacteremia with [piperacillin/tazobactam], that is an absolute no-go.* [Expert 4, Clinical Center 1]

Second, the evaluation of the recommendations is dependent upon the level of professional experience (quote 3). Accordingly, the more advanced the professional experience, the more confident a recommendation was to be approved or rejected. However, the data indicate that the majority of respondents expressed consensus regarding the CDSS recommendation in each case study (Table 1).

Table . Assessments of the experts surveyed on the feasibility of the pilot version of the artificial intelligence–based clinical decision support system (N=19).

	CDSS recommendation for initial treatment		CDSS recommendation for re-evaluation	
	Agreeing with the recommendation, n (%)	Rejecting the recommendation, n (%)	Agreeing with the recommendation, n (%)	Rejecting the recommendation, n (%)
Use case 1	14 (78)	4 (22)	12 (67)	6 (33)
Use case 2	11 (61)	7 (39)	1 (5)	18 (95)
Use case 3	3 (16)	16 (84)	2 (11)	16 (89)
Use case 4	15 (83)	3 (17)	14 (78)	4 (22)
Use case 5	13 (76)	4 (24)	12 (71)	5 (29)

A suggestion was made that the pilot CDSS should be expanded to include the integration of combination therapies. The vital parameters and patient information should be included as variables in the statistical model for predicting therapy to include the suspected focus of infection. In addition, it was noted that the data provided for the development of the model and the case studies in the pilot version were incomplete in many instances (Table 1). The physicians encountered significant challenges in evaluating the CDSS recommendation when the fields for data that could be viewed in the application were left empty (quote 4).

> ***Quote 4:****Here again the PCT [Procalcitonin] is missing, that would make it easier for us. We only have [...] indirect information that we have an infection here [...] measured by the thrombocytopenia, and then [...] I can get along with actually using piperacillin and tazobactam primarily here when it comes to the calculated antibiotic therapy [...].* [Expert 5, Clinical Center 3]

> ***Quote 5:****The [application] is more explanatory than my phone. So, everything is fine in that respect.* [Expert 3, Clinical Center 3]

From the physicians' perspective, the primary factor influencing the acceptance of the AI-based CDSS was the time required for completion of the process. If the system is able to produce an evident medication recommendation in a shorter time than the physicians can achieve, a high willingness to use it was reported. There was unanimous feedback that the design and layout of the CDSS are pleasant and easy to understand (quote 5). Furthermore, the majority of respondents indicated a preference for a mobile version of the system for use in everyday clinical practice. This preference was again justified by the availability of digital resources.

Facilitating and Inhibiting Factors for the Implementation of an AI-Based Clinical Decision Support Systems

The focus group interviews showed that while there is already strong foundational support for digital tools and applications in diagnostics and therapy as part of routine clinical practice, the use of more advanced applications, especially those based on AI, was limited to isolated cases and primarily on a project basis. Previous experiences with CDSSs had a significant impact on expectations regarding the quality of the CDSS, the effort required for integration into everyday medical practice, and the intention to potentially use such a system (quotes 6 and 7).

> ***Quote 6:****I'm thinking about the echocardiography, [...] where the AI has become so good that we let it*

take over. We do the examination ourselves. And yet the module helps us with that. That's how I imagine it [here] too. [Expert 2, focus group 2]

Quote 7:*Yes, for the colonoscopy [...] we once had [a CDSS]. But the human was faster than the AI. So in that respect [...] rather not.* [Expert 3, focus group 3]

Nevertheless, the majority of respondents indicated a high level of willingness to adopt AI-based CDSSs (quotes 8 and 9). The option of re-evaluation included in the application was considered to be particularly beneficial.

Quote 8:*I actually found the idea compelling. Because in the early days of my clinical work, even more so than now, the question always arises: am I doing it right now?* [Expert 1, focus group 2]

Quote 9:*[T]he field is so complex [...] that's why I think a support tool that works well is very helpful.* [Expert 3, focus group 1]

As previously noted, the acceptance and implementation of digital solutions is largely dependent on the amount of time available. In addition, the majority of respondents indicated that they are consistently seeking tools that facilitate efficient support in routine clinical practice. Consequently, the respondents exhibited an intrinsic motivation to use CDSSs, which was observed across all age groups and genders, but was particularly pronounced in the "30 - 40 years" and "40 - 50 years" age groups.

In contrast with the physician's technology openness as a beneficial factor, the level of digitization in the clinics was found to be heterogeneous. Furthermore, the physicians' assessment indicated that there was significant potential for improvement in all 3 hospitals (quote 10). The statements indicated that the level of digital infrastructure, including stable internet coverage, exhibited considerable variation not only between the clinics but also between individual wards within the clinics. Overall, there was a consensus that better equipment in terms of digital devices and applications was needed (quote 11). The lack of comprehensive digital and cross-sectoral availability of relevant patient health information was also identified as a barrier to the implementation of AI-based CDSSs and efficient medical treatment in general. This demonstrated that the degree of digitization in everyday clinical practice affects the perceived feasibility of implementing the CDSS and is the most significant barrier to the beneficial factors.

Quote 10:*So, it's really funny that you can still secure good pens. Because we do so much paper based.* [Expert 2, focus group 2]

Quote 11:*The problem is that there are far too many individual components that all [...] communicate with each other via interfaces. And in the process, values are lost or the data transfer doesn't work. And then you end up standing there and yes, the computer hangs.* [Expert 4, focus group 2]

Furthermore, the existing data basis for the AI-based CDSS model was identified as a significant challenge. The participants were aware that, in addition to the aforementioned lack of digitized patient information, the heterogeneous quality of previous antibiotic administration practice also had an influence on the predictive ability of the recommendation from the CDSS. Furthermore, an increased documentation and justification effort in the event of potential deviation from the CDSS recommendation was also seen as a barrier. However, all participants were aware that such a system would be implemented as a supportive and not a replacement measure, and this was not criticized.

Two relevant recommendations for improvement were derived from the interviews. First, the majority of respondents indicated that applicable guidelines on antibiotic therapy should be incorporated into the specification of the underlying statistical models. Second, there was a recurring suggestion that, alongside the further development of the presented system, the capability to diagnose sepsis infections should also be integrated into a CDSS. Once again, the time to correct diagnosis and subsequent appropriate treatment was identified as the main driver. Suggestions for adapting the tested CDSS comprised the inclusion of quantities and suggestions for the duration of antibiotic administration, in addition to the therapeutic agents themselves (quote 12).

Quote 12:*So, in an ideal world, [...] [AI] can be a huge step forward as a decision-making tool. But [A] the right information has to flow in. And [B] the decisive recommendation [...] is already anticipated here [...]. So, the first step: can this be sepsis? That is an important step. [...] In the second step, a recommendation for initial therapy is helpful [...], yes.* [Expert 4, focus group 1]

Discussion

Principal Findings

Overall, this use case analysis shows various barriers and facilitators for the implementation of AI-based CDSSs for antibiotic therapy in sepsis (Textbox 2). The findings of the expert and focus group interviews indicate that the potential of digitization and, in particular, AI-based tools is expected and viewed in a favorable light by the majority of the physicians. This finding is in line with other research which indicates that clinicians have a predominantly positive perception of AI systems [38-40]. Similarly, the willingness to use these tools is high [41]. In this context, the plausibility of the prediction and the potential time savings are of particular importance. It is evident that the existing contextual conditions represent the most significant obstacle to future implementation. In addition to the basic technical equipment in the clinics, other factors, such as a stable, comprehensive internet supply and the necessary digital availability of data, were also identified as potential barriers to future implementation. Accordingly, the question arises as to what extent clinics are equipped with the digital infrastructure to fully leverage the benefits of AI-based CDSSs. At this point it has to be noted that the digital maturity of German hospitals appears to be comparatively limited [42]. This can be attributed, at least in part, to a lack of financial resources [24,43]. It remains to be seen to what extent hospitals use the transformation fund provided by the German Hospital

Future Act and how this can enhance the conditions for the implementation of AI systems.

Textbox 2. Summary of barriers, facilitators, and recommendations for artificial intelligence–based clinical decision support systems' feasibility and implementation.

Barriers

- Overall
- Low degree of digitization in everyday clinical practice.
- Lack of comprehensive digital and cross-sectoral data availability.
- Negative previous experiences with clinical decision support systems.
- Use Cases
- Heterogeneous quality of previous antibiotic administration practice limits predictive ability of the artificial intelligence–based recommendations.
- Increased effort to document and justify potential deviations from the clinical decision support system's recommendation.
- The confidence to approve or reject artificial intelligence–based antibiotic recommendations depends on the level of professional experience.

Facilitators

- Overall
- Potential time savings.
- Physician's technology openness or intrinsic motivation.
- Positive previous experiences with clinical decision support systems.
- Use Cases
- Re-evaluation of clinical decision support system recommendations is considered to be particularly beneficial.
- Pleasant design and layout of the clinical decision support system (easy to understand).

Recommendations

- Incorporation of guidelines on antibiotic therapy into the specification of the statistical models.
- Clinical decision support systems should include the integration of combination therapies.
- Statistical model for predicting antibiotic recommendations should include vital parameters and further patient information to identify the suspected focus of infection.
- Besides sepsis treatment, the capability to diagnose sepsis infections should also be integrated into a clinical decision support system.
- Preference for a mobile version of the clinical decision support system.

The medication recommendations by the CDSS were not yet perceived as reliable due to the limited and biased data available for training the AI system. Often, a very broad standard therapy (piperacillin/tazobactam) was recommended. This was mainly due to the fact that this medication occurred considerably more frequently within the learning data set than all other antibiotics. According to the experts, this approach is not absolutely mandatory and sometimes maybe counterproductive in terms of calculated antibiotic therapy. Younger physicians in particular tended to agree with the recommendations. This is a sign that a CDSS for the most common "standard medication" might not be beneficial if not detrimental. This can be confirmed by the physicians' statements that they would like decision support primarily for medication combinations and for rare antibiotic agents or rare sources of infection. In view of the goal of avoiding the development of resistance by reducing the use of broad-spectrum antibiotics, the current pilot version of the AI system would possibly even promote this risk of further resistance. To avoid this, current guidelines should be incorporated into the model design in addition to retrospective primary data.

The respondents' call for the upstream integration of sepsis diagnostics in addition to the further development of the existing CDSS appears to be purposeful. A recent study has confirmed initial successes in this regard [44]. Although no clear correlations between sociodemographic characteristics and the willingness to use CDSSs were found in the present evaluation, differences in the approval or rejection of the AI-based CDSS recommendations depending on professional experience were identified. This finding is supported by Lambert et al [45]. It is recommended that future studies investigate the degree to which professional experience and other sociodemographic variables exert an influence on technology acceptance and the willingness to use CDSSs. In this regard, using the UTAUT would be appreciated [35].

Generally, the following challenges associated with the use of AI in health care should not be overlooked. For example, automation bias, in which health care providers may develop an overreliance on AI recommendations, potentially reducing their clinical vigilance and leading to oversights in patient care, needs to be considered [46]. Furthermore, AI-driven CDSSs may unintentionally propagate biases present in training datasets, resulting in discriminatory recommendations that may disadvantage certain demographic groups [47]. Another key concern is the ambiguity and complexity of liability; it remains unclear who is legally responsible if an AI-influenced decision contributes to a negative patient outcome [48]. Furthermore, previous research has highlighted the potential negative impact of a lack of user involvement in the development of CDSSs on the willingness to use them, even when the system quality is optimal [26]. One of the key strengths of the study is the early involvement of future users. The participatory approach highlights the need for involvement. Consequently, given the significant lack of engagement with CDSS interventions reported within the literature [49], future studies should consider the involvement of users from the earliest stages of development [26,28,50]. The findings of the present study confirm that in addition to personal factors, the plausibility of the recommendations is the most decisive factor determining the willingness to use the tool [51]. Further development of the pilot version or similar systems is therefore crucial for implementing and exploiting the potential benefits of AI. Furthermore, the feedback from the physicians participating in this study provided valuable insights into additional variables that were subsequently integrated into the further development of the CDSS, following the completion of the interviews.

Limitations

Several limitations must be acknowledged in the evaluation of the use case presented in this study. First, the efficacy of the CDSS was clearly limited due to the rather small amount of data used for training. The findings suggest that the CDSS may offer less benefit for physicians during the initial therapy phase compared with its use in the context of therapy re-evaluation. At this stage, it was not feasible to accurately predict the most appropriate antibiotic, one that would minimize treatment changes and side effects, based on the currently available data. The acceptance of AI-based interventions heavily depends on the plausibility and relevance of therapy recommendations, which ultimately impacts their efficacy. Therefore, further development and refinement of AI-based CDSSs are essential, along with an expansion of the digital data infrastructure.

In addition, it is important to consider that the evaluation was based on a qualitative interview study, which comes with inherent limitations. The voluntary nature of participation may introduce self-selection bias, potentially leading to an overestimation of the participants' openness to technology and their acceptance of AI-based CDSSs. This limitation suggests that the findings may not be fully generalizable to the broader population of health care professionals.

Conclusion

This pilot study of an AI-based CDSS indicates that further development is needed to achieve the original goals of minimizing switches in antibiotic prescriptions and reducing the reliance on broad-spectrum antibiotics. Relying solely on retrospective data from past care practices does not seem to be an effective strategy for meeting these objectives. In addition, it is essential to address the key challenges identified in this study to enable the successful integration of AI into clinical settings. While the respondents generally showed a positive attitude toward the use of AI-based CDSSs, a more comprehensive evaluation of technology acceptance among health care professionals can only be conducted when a version with realistic, guideline-compliant recommendations is available. Early engagement of users in the development process proves beneficial, not only for refining the CDSS but also for fostering acceptance among medical practitioners.

Acknowledgments

The authors would like to thank all members of the KINBIOTICS project and in particular the contact persons at the 3 hospitals (Evangelisches Klinikum Bethel, Klinikum Bielefeld Mitte, and Klinikum Lippe) for their support in conducting the interviews. Special thanks go to all the physicians who took part in the interviews and to Inga Heißenberg, who helped to prepare and conduct the interviews. Furthermore, we would like to thank Dr Timothy Mc Call for his support in translating the original quotations. The KINBIOTICS project was funded by the German Federal Ministry of Health (grant ZMVI1-2520DAT930). The funding organization was not involved in the data collection, evaluation, analysis, or writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Standards for Reporting Qualitative Research checklist.
[PDF File, 121 KB - humanfactors_v12i1e66699_app1.pdf]

References

1. Kumar NR, Balraj TA, Kempegowda SN, Prashant A. Multidrug-resistant sepsis: a critical healthcare challenge. Antibiotics (Basel) 2024 Jan 4;13(1):46. [doi: 10.3390/antibiotics13010046] [Medline: 38247605]

2. GBD 2021 Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance 1990-2021: a systematic analysis with forecasts to 2050. Lancet 2024 Sep 28;404(10459):1199-1226. [doi: 10.1016/S0140-6736(24)01867-1] [Medline: 39299261]
3. Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. Lancet 2020 Jan 18;395(10219):200-211. [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31954465]
4. Fleischmann-Struzek C, Mikolajetz A, Schwarzkopf D, et al. Challenges in assessing the burden of sepsis and understanding the inequalities of sepsis outcomes between National Health Systems: secular trends in sepsis and infection incidence and mortality in Germany. Intensive Care Med 2018 Nov;44(11):1826-1835. [doi: 10.1007/s00134-018-5377-4] [Medline: 30284637]
5. Bauer M, Groesdonk HV, Preissing F, Dickmann P, Vogelmann T, Gerlach H. Mortality in sepsis and septic shock in Germany. Results of a systematic review and meta-analysis. Anaesthesist 2021 Aug;70(8):673-680. [doi: 10.1007/s00101-021-00917-8] [Medline: 33559687]
6. The burden of antimicrobial resistance in G7 countries and globally: an urgent call for action. Robert Koch-Institut. 2022. URL: https://edoc.rki.de/handle/176904/9864;jsessionid=BB6B616866626A2E3B56758E5A1C5C8B?locale-attribute=en [accessed 2025-02-15]
7. Rödenbeck M, Ayobami O, Eckmanns T, Pletz MW, Bleidorn J, Markwart R. Clinical epidemiology and case fatality due to antimicrobial resistance in Germany: a systematic review and meta-analysis, 1 January 2010 to 31 December 2021. Euro Surveill 2023 May;28(20):2200672. [doi: 10.2807/1560-7917.ES.2023.28.20.2200672] [Medline: 37199987]
8. Zilberberg MD, Shorr AF, Micek ST, Vazquez-Guillamet C, Kollef MH. Multi-drug resistance, inappropriate initial antibiotic therapy and mortality in Gram-negative severe sepsis and septic shock: a retrospective cohort study. Crit Care 2014 Nov 21;18(6):596. [doi: 10.1186/s13054-014-0596-8] [Medline: 25412897]
9. Pradipta IS, Sodik DC, Lestari K, et al. Antibiotic resistance in sepsis patients: evaluation and recommendation of antibiotic use. N Am J Med Sci 2013 Jun;5(6):344-352. [doi: 10.4103/1947-2714.114165] [Medline: 23923107]
10. Kalın G, Alp E, Chouaikhi A, Roger C. Antimicrobial multidrug resistance: clinical implications for infection management in critically ill patients. Microorganisms 2023 Oct 16;11(10):2575. [doi: 10.3390/microorganisms11102575] [Medline: 37894233]
11. Neugebauer M, Ebert M, Vogelmann R. A clinical decision support system improves antibiotic therapy for upper urinary tract infection in a randomized single-blinded study. BMC Health Serv Res 2020 Mar 6;20(1):185. [doi: 10.1186/s12913-020-5045-6] [Medline: 32143630]
12. Van Dort BA, Penm J, Ritchie A, Baysari MT. The impact of digital interventions on antimicrobial stewardship in hospitals: a qualitative synthesis of systematic reviews. J Antimicrob Chemother 2022 Jun 29;77(7):1828-1837. [doi: 10.1093/jac/dkac112] [Medline: 35388414]
13. Xu R, Wu L, Wu L, Xu C, Mu T. Effectiveness of decision support tools on reducing antibiotic use for respiratory tract infections: a systematic review and meta-analysis. Front Pharmacol 2023;14(1253520):1253520. [doi: 10.3389/fphar.2023.1253520] [Medline: 37745052]
14. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3(17):17. [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
15. Graafsma J, Murphy RM, van de Garde EMW, et al. The use of artificial intelligence to optimize medication alerts generated by clinical decision support systems: a scoping review. J Am Med Inform Assoc 2024 May 20;31(6):1411-1422. [doi: 10.1093/jamia/ocae076] [Medline: 38641410]
16. Rittmann B, Stevens MP. Clinical decision support systems and their role in antibiotic stewardship: a systematic review. Curr Infect Dis Rep 2019 Jul 24;21(8):29. [doi: 10.1007/s11908-019-0683-8] [Medline: 31342180]
17. Rutten JJS, van Buul LW, Smalbrugge M, et al. An electronic health record integrated decision tool and supportive interventions to improve antibiotic prescribing for urinary tract infections in nursing homes: a cluster randomized controlled trial. J Am Med Dir Assoc 2022 Mar;23(3):387-393. [doi: 10.1016/j.jamda.2021.11.010] [Medline: 34896069]
18. Hartman EAR, Groen WG, Heltveit-Olsen SR, et al. Implementation of a tailored multifaceted antibiotic stewardship intervention to improve antibiotic prescribing for urinary tract infections in frail older adults (ImpresU) in four European countries: a process evaluation alongside a pragmatic cluster randomized controlled trial. Trials 2024 Oct 18;25(1):691. [doi: 10.1186/s13063-024-08545-4] [Medline: 39425170]
19. Laka M, Milazzo A, Merlin T. Can evidence-based decision support tools transform antibiotic management? A systematic review and meta-analyses. J Antimicrob Chemother 2020 May 1;75(5):1099-1111. [doi: 10.1093/jac/dkz543] [Medline: 31960021]
20. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. BMJ 2005 Apr 2;330(7494):765. [doi: 10.1136/bmj.38398.500764.8F] [Medline: 15767266]

21. Catho G, Sauser J, Coray V, et al. Impact of interactive computerised decision support for hospital antibiotic use (COMPASS): an open-label, cluster-randomised trial in three Swiss hospitals. Lancet Infect Dis 2022 Oct;22(10):1493-1502. [doi: 10.1016/S1473-3099(22)00308-5] [Medline: 35870478]
22. Hong N, Liu C, Gao J, et al. State of the art of machine learning-enabled clinical decision support in intensive care units: literature review. JMIR Med Inform 2022 Mar 3;10(3):e28781. [doi: 10.2196/28781] [Medline: 35238790]
23. Komorowski M, Celi LA, Badawi O, Gordon AC, Faisal AA. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nat Med 2018 Nov;24(11):1716-1720. [doi: 10.1038/s41591-018-0213-5] [Medline: 30349085]
24. Weinert L, Müller J, Svensson L, Heinze O. Perspective of information technology decision makers on factors influencing adoption and implementation of artificial intelligence technologies in 40 German hospitals: descriptive analysis. JMIR Med Inform 2022 Jun 15;10(6):e34678. [doi: 10.2196/34678] [Medline: 35704378]
25. KINBIOTICS [Website in German]. URL: https://kinbiotics.de [accessed 2024-11-07]
26. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24. [doi: 10.2196/medinform.8912] [Medline: 29669706]
27. Ji M, Yu G, Xi H, Xu T, Qin Y. Measures of success of computerized clinical decision support systems: an overview of systematic reviews. Health Policy Technol 2021 Mar;10(1):196-208. [doi: 10.1016/j.hlpt.2020.11.001]
28. Carroll C, Marsden P, Soden P, Naylor E, New J, Dornan T. Involving users in the design and usability evaluation of a clinical decision support system. Comput Methods Programs Biomed 2002 Aug;69(2):123-135. [doi: 10.1016/s0169-2607(02)00036-6] [Medline: 12100792]
29. Newton N, Bamgboje-Ayodele A, Forsyth R, Tariq A, Baysari MT. Does involving clinicians in decision support development facilitate system use over time? a systematic review. Stud Health Technol Inform 2023 Jun 22;304:11-15. [doi: 10.3233/SHTI230359] [Medline: 37347561]
30. Westerbeek L, Ploegmakers KJ, de Bruijn GJ, et al. Barriers and facilitators influencing medication-related CDSS acceptance according to clinicians: a systematic review. Int J Med Inform 2021 Aug;152:104506. [doi: 10.1016/j.ijmedinf.2021.104506] [Medline: 34091146]
31. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251. [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
32. Döring N. Datenanalyse. In: Döring N, editor. Forschungsmethoden Und Evaluation in Den Sozial- Und Humanwissenschaften [Book in German]: Berlin, Heidelberg: Springer Berlin Heidelberg; 2023:587-766. [doi: 10.1007/978-3-662-64762-2_12]
33. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
34. Venkatesh V, Thong J, Xu X, University of Arkansas, Hong Kong University of Science and Technology, The Hong Kong Polytechnic University. Unified theory of acceptance and use of technology: a synthesis and the road ahead. JAIS 2016;17(5):328-376. [doi: 10.17705/1jais.00428]
35. Dingel J, Kleine AK, Cecil J, Sigl AL, Lermer E, Gaube S. Predictors of health care practitioners' intention to use AI-enabled clinical decision support systems: meta-analysis based on the unified theory of acceptance and use of technology. J Med Internet Res 2024 Aug 5;26:e57224. [doi: 10.2196/57224] [Medline: 39102675]
36. Kuckartz U. Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung 4. Auflage. Weinheim [Book in German]: Basel: Beltz Juventa; 2018.
37. MAXQDA. URL: https://www.maxqda.com/ [accessed 2025-02-20]
38. van der Meijden SL, de Hond AAH, Thoral PJ, et al. Intensive care unit physicians' perspectives on artificial intelligence-based clinical decision support tools: preimplementation survey study. JMIR Hum Factors 2023 Jan 5;10:e39114. [doi: 10.2196/39114] [Medline: 36602843]
39. Maassen O, Fritsch S, Palm J, et al. Future medical artificial intelligence application requirements and expectations of physicians in German university hospitals: web-based survey. J Med Internet Res 2021 Mar 5;23(3):e26646. [doi: 10.2196/26646] [Medline: 33666563]
40. Pecqueux M, Riediger C, Distler M, et al. The use and future perspective of artificial Intelligence-a survey among German surgeons. Front Public Health 2022;10:982335. [doi: 10.3389/fpubh.2022.982335] [Medline: 36276381]
41. Hennrich J, Kauffmann AL, Buck C, Eymann T. Künstliche intelligenz in der hausärztlichen versorgung. In: Pfannstiel MA, editor. Künstliche Intelligenz Im Gesundheitswesen: Entwicklungen, Beispiele Und Perspektiven [Book in German] 2022:413-430. [doi: 10.1007/978-3-658-33597-7_18]
42. Geissler A, Hollenbach J, Haring M, Amelung VE, Thun S, Haering A. A nationwide digital maturity assessment of hospitals – results from the German DigitalRadar. Health Policy Technol 2024 Sep;13(4):100904. [doi: 10.1016/j.hlpt.2024.100904]
43. Burmann A, Fischer B, Brinkkötter N, Meister S. Managing directors' perspectives on digital maturity in German hospitals-a multi-point online-based survey study. Int J Environ Res Public Health 2022 Aug 6;19(15):9709. [doi: 10.3390/ijerph19159709] [Medline: 35955066]
44. Steinbach D, Ahrens PC, Schmidt M, et al. Applying machine learning to blood count data predicts sepsis with ICU admission. Clin Chem 2024 Mar 2;70(3):506-515. [doi: 10.1093/clinchem/hvae001] [Medline: 38431275]

45. Lambert SI, Madi M, Sopka S, et al. An integrative review on the acceptance of artificial intelligence among healthcare professionals in hospitals. NPJ Digit Med 2023 Jun 10;6(1):111. [doi: 10.1038/s41746-023-00852-5] [Medline: 37301946]
46. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc 2012;19(1):121-127. [doi: 10.1136/amiajnl-2011-000089] [Medline: 21685142]
47. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA Intern Med 2018 Nov 1;178(11):1544-1547. [doi: 10.1001/jamainternmed.2018.3763] [Medline: 30128552]
48. Price WN 2nd, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA 2019 Nov 12;322(18):1765-1766. [doi: 10.1001/jama.2019.15064] [Medline: 31584609]
49. Rawson TM, Moore LSP, Hernandez B, et al. A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately? Clin Microbiol Infect 2017 Aug;23(8):524-532. [doi: 10.1016/j.cmi.2017.02.028] [Medline: 28268133]
50. Neff MC, Schaaf J, Noll R, et al. Initial user-centred design of an AI-based clinical decision support system for primary care. Stud Health Technol Inform 2024 Jan 25;310:1051-1055. [doi: 10.3233/SHTI231125] [Medline: 38269975]
51. Panicker RO, George AE. Adoption of automated clinical decision support system: a recent literature review and a case study. Archives of Medicine and Health Sciences 2023;11(1):86-95. [doi: 10.4103/amhs.amhs_257_22]

Abbreviations

AI: artificial intelligence
AMR: antimicrobial resistance
CDSS: clinical decision support system
ICU: intensive care unit
SRQR: Standards for Reporting Qualitative Research
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk; submitted 02.10.24; peer-reviewed by D Sharma, GD Giebel, T Bommhardt; revised version received 22.11.24; accepted 31.12.24; published 04.03.25.

Please cite as:
Düvel JA, Lampe D, Kirchner M, Elkenkamp S, Cimiano P, Düsing C, Marchi H, Schmiegel S, Fuchs C, Claßen S, Meier KL, Borgstedt R, Rehberg S, Greiner W
An AI-Based Clinical Decision Support System for Antibiotic Therapy in Sepsis (KINBIOTICS): Use Case Analysis
JMIR Hum Factors 2025;12:e66699
URL: https://humanfactors.jmir.org/2025/1/e66699
doi:10.2196/66699

The Impact of Human-Robot Collaboration Levels on Postural Stability During Working Tasks Performed While Standing: Experimental Study

Daniele Bibbo, PhD; Giovanni Corvini, PhD; Maurizio Schmid, PhD; Simone Ranaldi, PhD; Silvia Conforto, PhD

Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Via Vito Volterra, 62 - Corpo B, Roma, Italy

Corresponding Author:
Daniele Bibbo, PhD
Department of Industrial, Electronic and Mechanical Engineering, Roma Tre University, Via Vito Volterra, 62 - Corpo B, Roma, Italy

Abstract

Background: The integration of collaborative robots (cobots) in industrial settings has the potential to enhance worker safety and efficiency by improving postural control and reducing biomechanical risk. Understanding the specific impacts of varying levels of human-robot collaboration on these factors is crucial for optimizing cobot use.

Objective: This study aims to investigate the biomechanical effects of different levels of human-robot collaboration on postural stability and control during simulated working tasks.

Methods: A total of 14 participants performed simulated cashier working activities under 4 different collaboration modalities, with increasing levels of cobot assistance: full (Fu), half robot touch (HRT), half robot (HRb), and full robot (FRb). Center of pressure trajectories were extracted from 2 force plates' data to calculate 4 posturography parameters—mean distance (MDIST), mean velocity (MVELO), 95% confidence ellipse area (AREA-CE), and sway area (AREA-SW)—which were analyzed to assess the impact of cobot intervention on postural control.

Results: Nonparametric tests showed significance in the effect of the collaboration modalities on the 4 analyzed parameters. Post hoc tests revealed that FRb modality led to the greatest enhancement in postural stability, with a reduction in MDIST (4.2, SD 1.3 cm in Fu vs 1.6, SD 0.5 cm in FRb) and MVELO (16.3, SD 5.2 cm/s in Fu vs 7.9, SD 1.1 cm/s in FRb). AREA-CE and AREA-SW also decreased significantly with higher levels of cobot assistance (AREA-CE: 134, SD 91 cm^2 in Fu vs 22, SD 12 cm^2 in FRb; AREA-SW: 16.2, SD 8.4 cm^2/s in Fu vs 4.0, SD 1.6 cm^2/s in FRb). Complete assistance of the cobot significantly reduced interindividual variability of all center of pressure parameters. In FRb modality, as compared with all other conditions, removing the weight of the object during loading or unloading phases caused a significant decrease in all parameter values.

Conclusions: Increased cobot assistance significantly enhances postural stability and reduces biomechanical load on workers during simulated tasks. Full assistance from cobots, in particular, minimizes postural displacements, indicating more consistent postural control improvements across individuals. However, high levels of cobot intervention also reduced the natural variation in how people balanced themselves. This could potentially lead to discomfort in the long run. Midlevel cobot assistance modalities can thus be considered as a good compromise in reducing biomechanical risks associated with postural stability at the same time granting a satisfactory level of user control.

(JMIR Hum Factors 2025;12:e64892) doi:10.2196/64892

KEYWORDS

human-robot collaboration; cobot assistance; postural control; biomechanical risk; ergonomics; collaborative robot

Introduction

Industrial manufacturing is transitioning from well-established production procedures toward more flexible and intelligent manufacturing systems (Industry 4.0, as initially described by Kagermann et al [1]). This evolution aims to develop innovative, sustainable solutions that create new business models, improve working conditions, increase plant productivity, and enhance product quality [2]. Robotics plays a crucial role in this context, with industrial robots being widely adopted due to their ability to relieve humans from repetitive, unhealthy, or dangerous tasks [3]. The field of robotics has evolved to not only allow humans to share the same workspace with robots but also use them as assistants, thereby enhancing human-robot collaboration (HRC) [4].

Collaborative robots (cobots) are at the forefront of this revolution, offering increased productivity, flexibility, versatility, and safety compared with traditional industrial robots. Cobots are designed to work alongside humans, sharing the same workspace without the need for safety cages and

providing greater mobility and flexibility [5]. The integration of cobots in industrial environments has been shown to significantly enhance operational efficiency [6] and worker satisfaction by enabling safer and more ergonomic working conditions [7,8]. However, the integration of cobots impacts various human factors, which include psychological aspects [9], emotions [10], and biomechanical effects [11]. Consequently, it is important to consider both cognitive ergonomics and biomechanical safety to optimize the overall effectiveness and well-being of human workers [12]. For instance, Gualtieri et al [13] have investigated the cognitive elements involved in human interaction with cobots, revealing significant insights into how mental workload and task complexity can affect worker performance and satisfaction [14]. Similarly, improper collaboration with cobots can expose workers to biomechanical risks, even though these robots are intended to alleviate heavy and tedious tasks. Previous research studies [15,16] emphasized the importance of ergonomic considerations in HRC to prevent musculoskeletal disorders (MSDs) among workers.

Thus, the right choice of cobot collaboration modality can reduce biomechanical overload on workers, ensuring their safety.

Recent advancements in sensor technology and data analytics have enabled detailed assessments of human biomechanics during HRC [17,18]. Studies have shown that real-time monitoring and analysis of physiological data can provide valuable insights into the physical strain experienced by workers and help in optimizing collaborative processes [19]. For example, the use of wearable sensors and motion capture systems could effectively measure and analyze the biomechanical load on workers during different collaborative tasks with robots [20]. Other recent studies investigated the effects of different HRC modalities on physiological human parameters such as trunk oscillations [21] and muscle coactivation [22,23]. Such studies can also be profitably taken into consideration for the optimization of workplace design and task allocation.

When work tasks are performed in quasi-static conditions, direct measures of posture could reveal insights into the stability and balance of workers and can thus be used to improve ergonomics during HRC, as they can help identify postural adjustments and biomechanical loads, providing a more comprehensive understanding of HRC ergonomic impact [24].

Following this principle, this study will evaluate and analyze posturography data to assess the biomechanical risks in HRC working contexts where cobots are used in different modalities. Through this evaluation, the study seeks to determine whether cobot assistance can effectively reduce biomechanical risks: by examining how different collaborative modalities influence workers' stability and balance, thus providing insights into the effects of cobot assistance on postural control, as this latter one indirectly reflects safety, efficiency, and ergonomic well-being of human workers in collaborative environments.

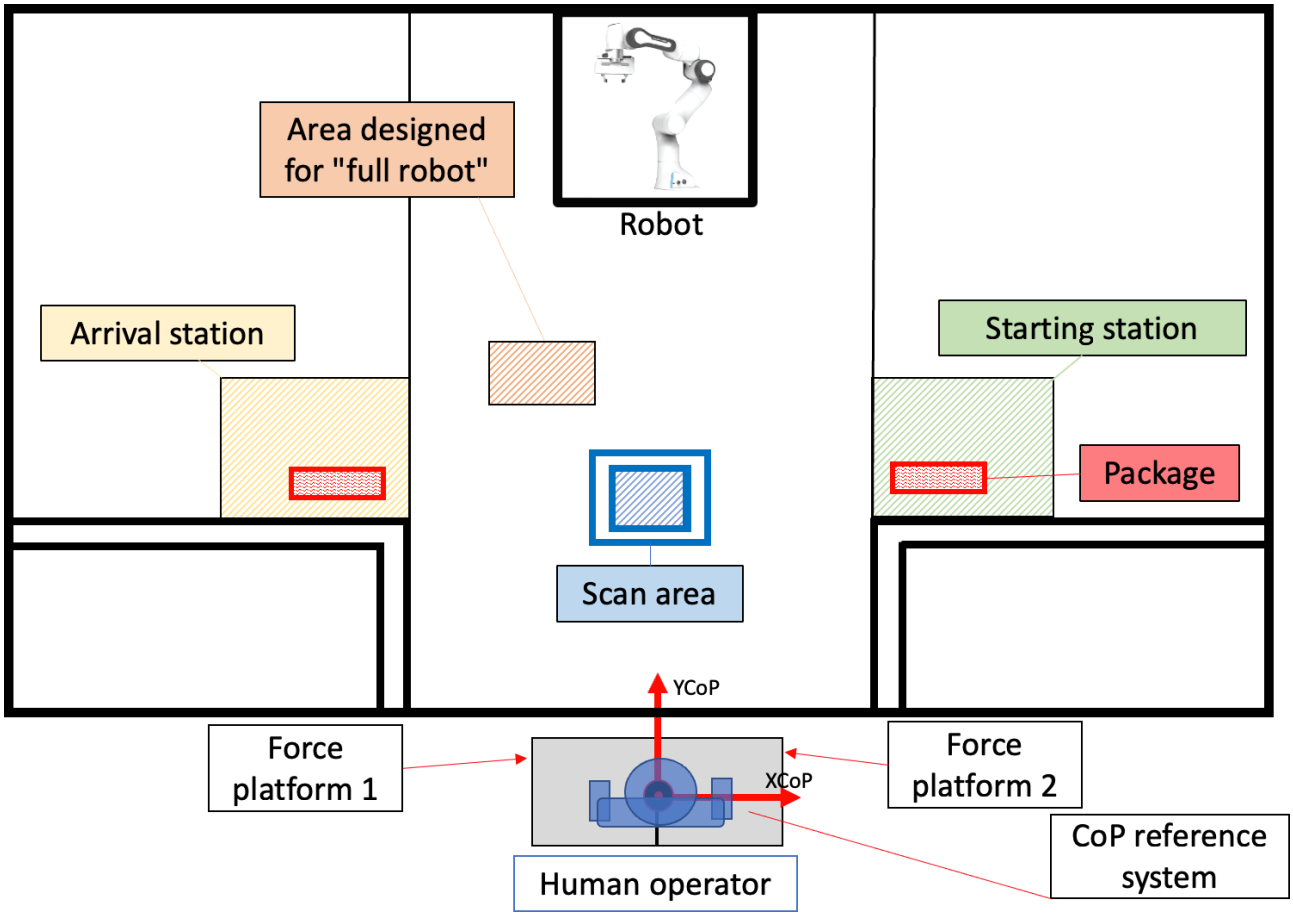
Methods

Experimental Setup

Overview

The setup was designed to compare a simulated work activity performed with and without cobot assistance. A specific workbench was created by positioning the cobot on one of the longer sides of a 2 × 1 m table, with the operator standing upright on the opposite side (Figure 1). The operator's standing position was fixed relative to the edge of the table, defined by the position of his or her feet and the projection onto the ground of the long side of the table. A starting station (ie, the initial position to the right of the operator) and an arrival station (ie, the final position to the left) were defined, each 1m apart from the other and 30 cm away from the operator in the direction of the cobot. A scanning area was defined in the center of the table in front of the operator. This workbench arrangement allowed the operator to perform all the activities (or part of them) independently, while the cobot could collaborate on some or none of the tasks without interfering with the operator's movements. This configuration was defined to minimize possible hindering of the operator's movements.

Figure 1. Experimental setup dimensioning for the simulation of load-moving activities on a workbench. CoP: center of pressure.



The activities involved the handling and relocating of rectangular packages, each measuring 14 × 12 × 5 cm and weighing 1 kg, bearing a QR code on one side.

The tests were conducted with a group of 14 male volunteers aged between 25 and 48 years (weight: 79, SD 7 kg; height: 180, SD 5 cm). All volunteers were healthy and had no evident or declared mobility issues. Moreover, given the design of the setup, only right-handed subjects have been involved in the tests. Participants were instructed to (1) grasp the packages from the starting station, (2) identify the QR code and scan it (ie, facing the QR code to the scan area) in the central area, and then (3) place the package in the arrival station. Different levels of collaboration were implemented by removing the first part of these activities (ie, grasping the package and bringing it to the center of the table) when in partial assistance, and removing also the last part (ie, bringing the package to the arrival station) when in full cobot assistance. Each participant was asked to perform the task as naturally as possible. This task was designed to simulate the work activity of a cashier in a supermarket, a highly relevant example in terms of strenuous activity [25], where an increase in the level of cobot assistance is hypothesized to reduce the associated biomechanical risk.

In the test, each participant performed consecutive repetitions of the same task, for a 5-minute duration, with different levels of cobot assistance.

The frequency of processing (packages per minute [PPM]) to be handled was not defined a priori, and participants were allowed to choose their own pace in performing all the activities. When collaborating with the cobot, the PPM frequency was however influenced by the cobot's package handling activity, as described below.

Each participant was asked to perform the described activity under 4 different collaboration modalities, performed in random order, with a 5-minute break between each test.

Full (Fu)

In this baseline scenario, the operator performs the tasks independently, without cobot collaboration. The operator picks up the package from the starting station with the right hand, examines it with both hands to locate the QR code, brings it close to the scanner simulation area, and finally places the package in the arrival station with the left hand. The processing rate was freely self-selected by each participant.

Half Robot Touch (HRT)

In this condition, the cobot task is to pick up the package from the starting station and wait for the operator to extend their right arm toward the same area, and to touch it to start. When touched, the cobot moves bringing the package in front of the operator who grabs and examines it with both hands to locate the QR code. The operator then brings it close to the scanner simulation area and finally places the package in the arrival station with the left hand. The task is kinematically like the "Full" case, but there is no lifting of the package from the starting station toward the scanning area. Moreover, the operator drives the activation of the cobot with a touch, and therefore the number of PPM is decided by everyone, taking into account the cobot speed.

Half Robot (HRb)

In this modality, the cobot's task is to pick up the package from the starting station and autonomously transport it in front of the operator, who grabs and examines it with both hands to locate the QR code. The operator then brings it close to the scanner simulation area and finally places the package in the arrival station with the left hand. The operator is not required to initially reach the right arm toward the starting station to ask the cobot for help, so the task is different in terms of kinematic features compared with the previous cases. Packages are moved continuously by the cobot while the operator is performing the required tasks, therefore, the cobot determines the number of PPM.

Full Robot (FRb)

In this condition, the cobot's task is to pick up the package from the starting station and autonomously transport it continuously in front of the operator, who grabs and examines it with both hands to locate the QR code. The operator then brings it close to the scanner simulation area and finally places the package in a closer area with respect to the arrival station (Figure 1). Meanwhile, the cobot picks up the package left by the operator and takes it to the arrival station. The cobot intervenes significantly, fully collaborating in the activity. However, the time taken by the cobot to move the packages always exceeds the time that the operator needs to perform the activity. Therefore, the operator must wait for the cobot, and the motor activity is discontinuous, with the number of PPM defined by the cobot kinematics.

Ethical Considerations

The experimental protocol was approved by the Commissione Etica dell'Università degli Studi Roma Tre (BRIC2019-BRISK approval r.01-10/06/2021). Informed consent was obtained from all participants prior to their inclusion in the study, and they were informed of their right to withdraw at any time without consequences. All data collected were anonymized to protect participants' privacy and confidentiality. No personally identifiable information was retained, and strict security measures were implemented to safeguard the data. Participants did not receive any financial or material compensation for their involvement in this study.

Data Acquisition and Processing

To assess the dynamic interaction of the operator with the ground, two 6-component force platforms (BTS P-6000) were positioned under the participants' feet. These platforms allow for measuring the ground reaction of each participant's lower limbs (force and moment components), from which the center of pressure (CoP) position of each foot is extracted during the execution of performed tasks. Data were acquired from force platforms using a sampling frequency of 500Hz (fixed by the HW system). An optoelectronic system (SMART DX 6000, 8 infrared cameras, 2 Mpixel@300 Hz) was employed to segment and identify each task repetition, and each phase of the repetition. Data from the optoelectronic system were acquired using a sampling frequency of 250Hz. Force platforms data were undersampled at 250 Hz to synchronize those with the optoelectronic system ones.

Dynamic Parameters Extraction

Trajectories of the CoP for each of the lower limbs (ie, CoP-right and CoP-left) during the execution of each task were extracted from the force plates data [26]. The total CoP is obtained as the weighted average between CoP-right and CoP-left and its displacement is defined by both the medial-lateral and the anterior-posterior coordinates with respect to a reference system centered between the 2 force plates (Figure 1).

Four CoP-derived parameters were calculated to evaluate the postural stability of the operator and quantify alterations in balance: for each of the collaboration modalities, the mean distance (MDIST), mean velocity (MVELO), 95% confidence ellipse area (AREA-CE), and sway area (AREA-SW) were calculated, according to the definitions reported in [27], for each cycle of the task repetition. For each parameter, mean values (as a measure of central tendency) and SD across repetitions (as a measure of intraindividual variability) were also calculated and underwent inferential statistics.

Cycles were identified using data extracted from motion capture data, which were synchronized with force plate data: in particular, the time location of the maximum excursion to the left in the medio-lateral direction of the marker placed on the right wrist was used to segment each cycle, as it corresponds to the instant when each repetition finishes.

Statistical Analysis

For each CoP-derived parameter thus obtained, the normality of the data, using the Shapiro-Wilk test, was assessed. When all distributions were normal, the homogeneity of the variances was checked with the Leneve test. If these conditions were not satisfied, the Kruskal-Wallis nonparametric test was applied to examine the effect of the collaboration modality. When significant, a post hoc test with a Bonferroni correction was conducted to investigate the differences among the 4 examined collaboration modalities. The α level was set to .05 and .01 (denoted by * and **, respectively).

Results

Rate of Package Processing

The average rate of package processing chosen by the operators in the absence of cobot assistance (Fu) was 15 (IQR 14-16) PPM. The different cobot collaboration modalities yielded median rates of 8 (IQR 7-8.5) for PPM for half robot touch (HRT), 12 (IQR 11.5-13) PPM for half robot (HRb), and 4 (IQR 3.9-4.1) PPM for full robot (FRb).

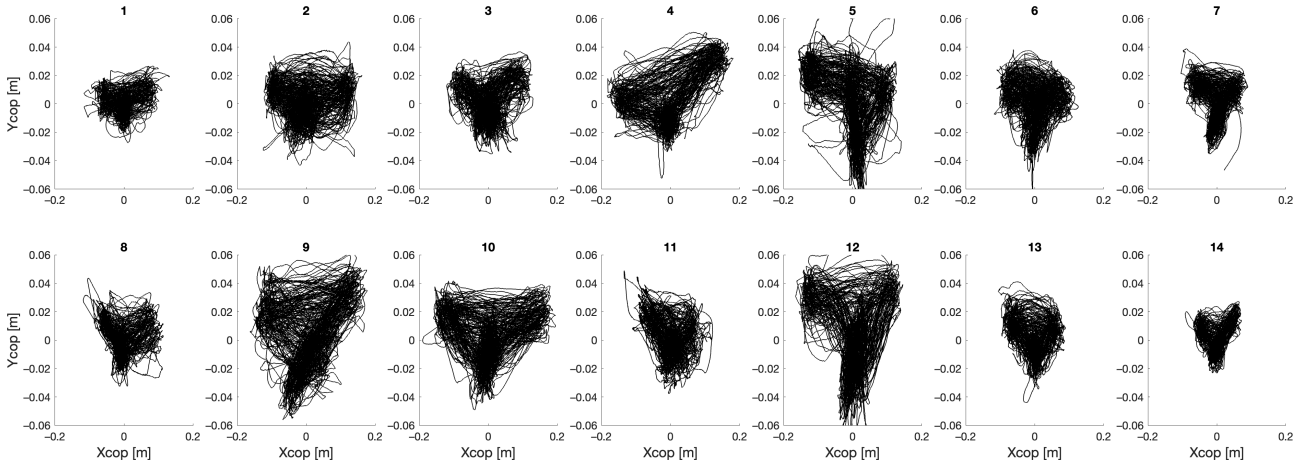
CoP Trajectories

In the following, the different trajectories of the CoP, for all 14 participants and for the 4 different collaboration modalities (ie, full [Fu], HRB, HRb, and FRb) are shown, with Xcop and Ycop denoting respectively the medium-lateral and the antero-posterior coordinate.

Considering the Fu modality (Figure 2), observed CoP paths can be classified into 2 main groups: participants (1,2,3,6,7,8,11,13,14) displayed a compact size of the CoP path, spread in both anterior-posterior (AP) and medial-lateral (ML)

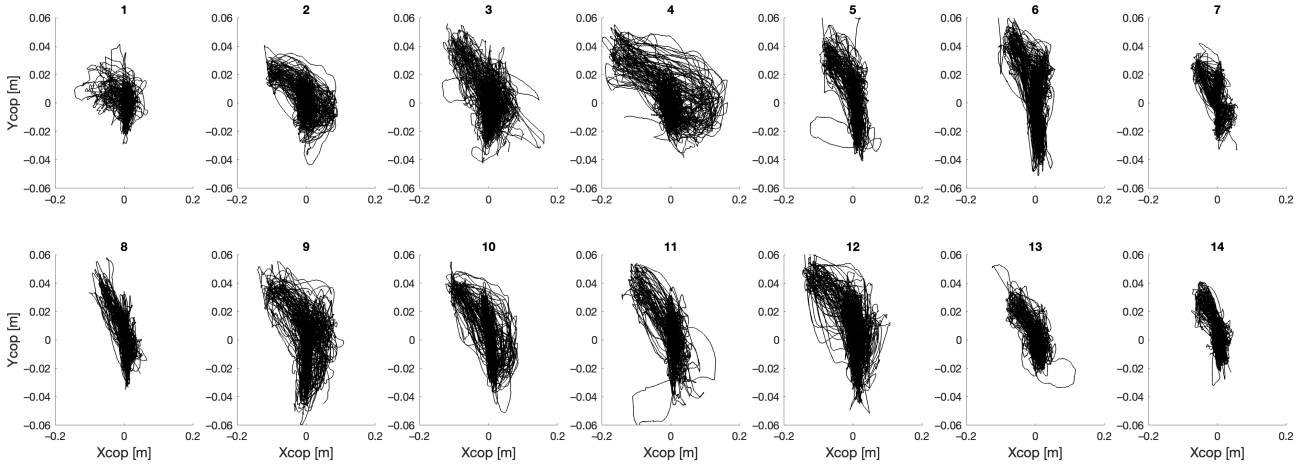
directions, while participants (4,5,9,10,12) drew wider areas, with a remarkable trajectory from the central backward position to the right forward position. However, all the CoP paths present quite diverse behavior, with no visible common aspects. Negative values for Ycop denote backward displacements.

Figure 2. Center of pressure (cop) trajectories for each participant during the full collaboration modality.



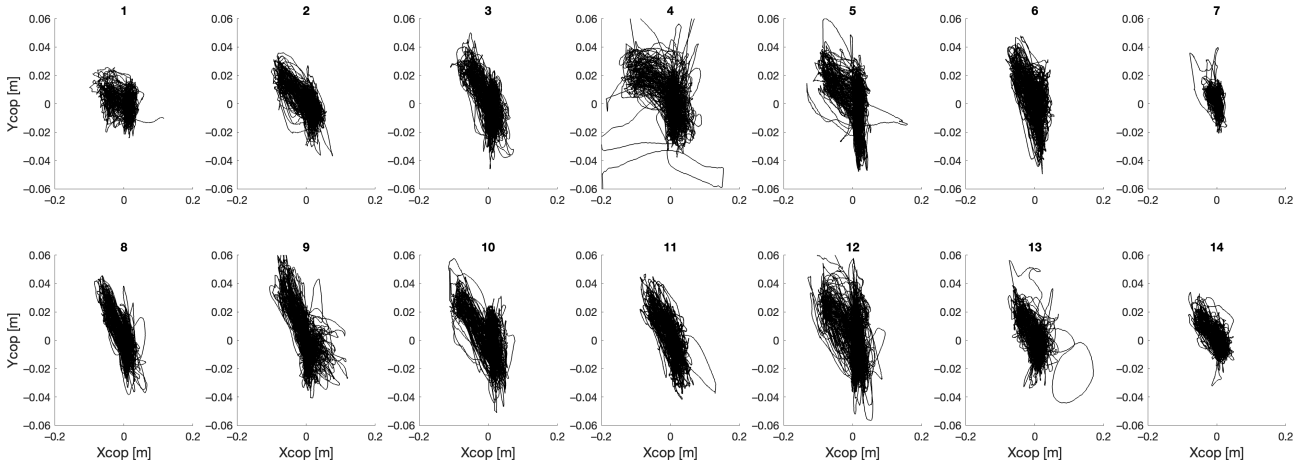
Considering the HRT collaboration modality (Figure 3), CoP path can be classified into 2 main groups also in this case: participants (1,2,7,8,13,14) maintained a compact size of the CoP path, more spread along the AP direction, while participants (3,4,5,6,9,10,11,12) drew a wider area of the CoP path. In this case, all CoP trajectories denoted a bias from the central backward position to the left forward position. Displayed CoP paths still exhibit rather diverse behavior, but with more commonalities as compared with Fu modality.

Figure 3. Center of pressure (cop) trajectories for each participant during the half robot touch collaboration modality.



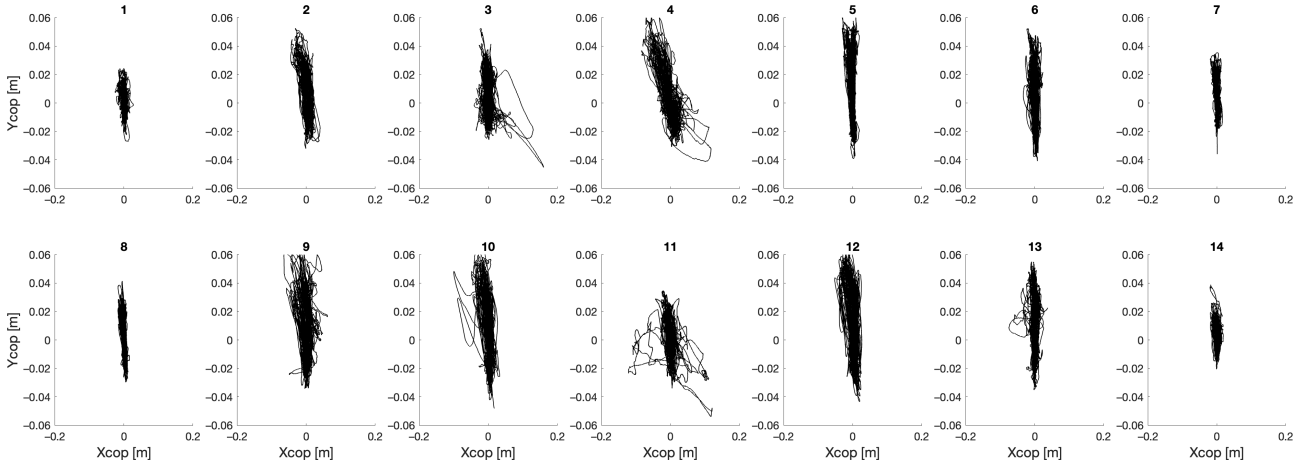
For the HRb collaboration modality (Figure 4), also in this case CoP path can be classified into 2 main groups: for participants (1,2,7,13,14) CoP path was maintained compact, differently from participants (3,4,5,6,8,9,10,11,12), especially considering the AP coordinate. Common behavior of CoP excursion across subjects was instead present along the ML coordinate, while the CoP bias from the central backward position to the left forward position observed in the HRT case is less evident.

Figure 4. Center of pressure (cop) trajectories for each participant under the half robot collaboration modality.



In the FRb collaboration modality (Figure 5), commonalities of CoP behavior are more relevant. While some participants still display lower excursions along the AP direction, the characteristics of CoP trajectories are mostly similar, with no bias both in the right and left forward positions. The amount of excursion in the ML direction is similar across participants.

Figure 5. Center of pressure (cop) trajectories for each participant during the full robot collaboration modality.



Normality Tests

The Shapiro-Wilk test indicated that some conditions of the parameters did not follow a normal distribution. For the MDIST values, a *P* value <.05, indicating rejection of the null hypothesis of normal distribution, was observed for the Fu condition (*P*=.048). In contrast, a normal distribution (*P* value >.05) was found for all the other conditions (HRT: *P*=.44; HRb: *P*=.57; FRb: *P*=.60).

Similar results were obtained for both the AREA-CE and AREA-SW parameters. The Fu condition showed a nonnormal distribution (*P*=.02 for AREA-CE and *P*=.04 for AREA-SW), while all other conditions were normally distributed (for AREA-CE: HRT, *P*=.18; HRb, *P*=.25; FRb, *P*=.33; for AREA-SW: HRT, *P*=.08; HRb, *P*=.86; FRb, *P*=.25).

Since at least 1 of the distributions for each parameter was not normal, nonparametric statistical analyses were applied.

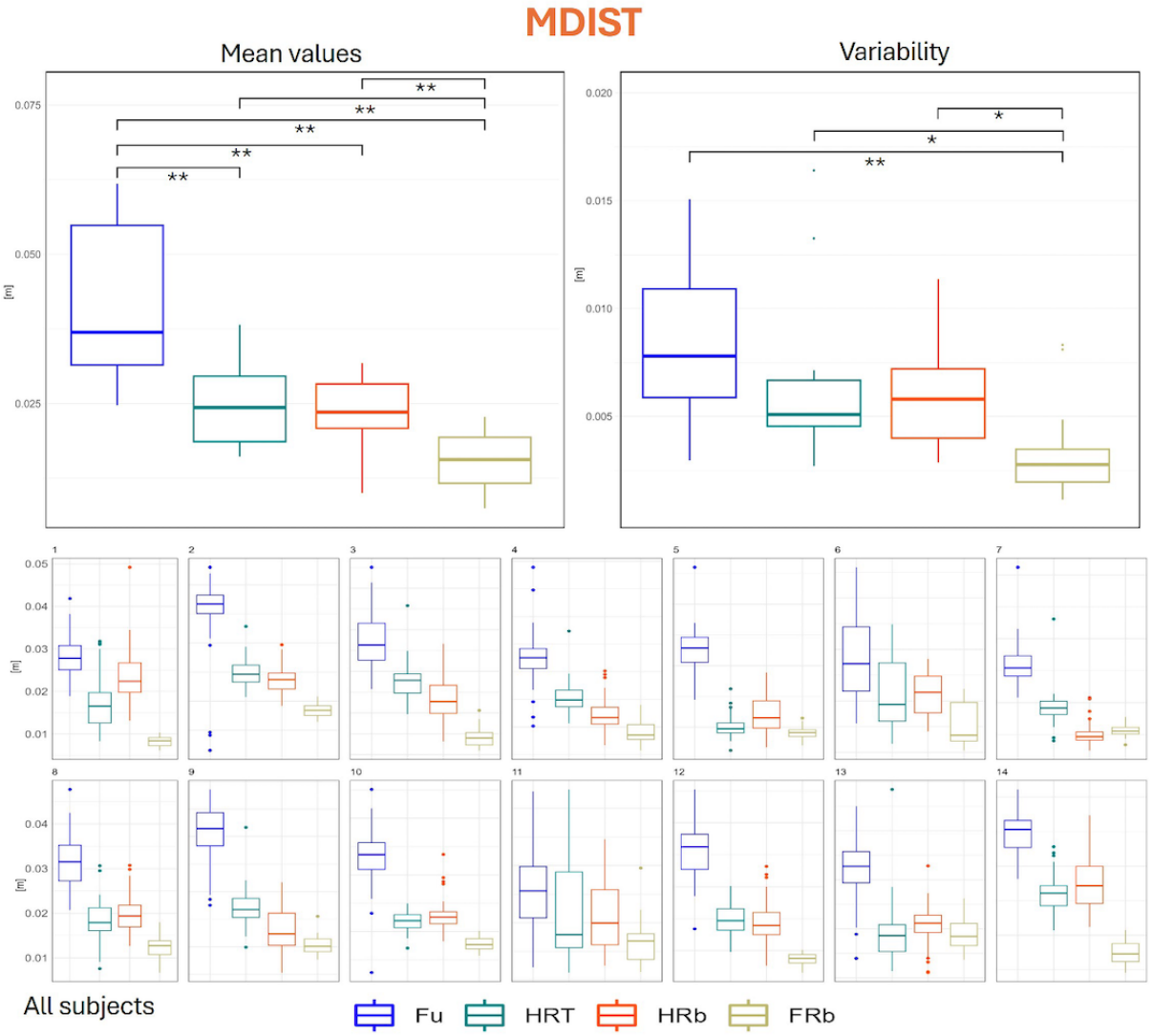
For the MVELO parameter, data were normally distributed across all conditions. However, Levene test revealed that the variances among the different conditions were not homogeneous (*P* value <.05). Consequently, a nonparametric test was also used for this parameter.

CoP-Derived Parameters

Descriptive statistics at the individual level, and descriptive and inferential statistics at the group level for MDIST, MVELO, AREA-CE, and AREA-SW, as obtained across the 4 collaboration modalities, are given in the following figures.

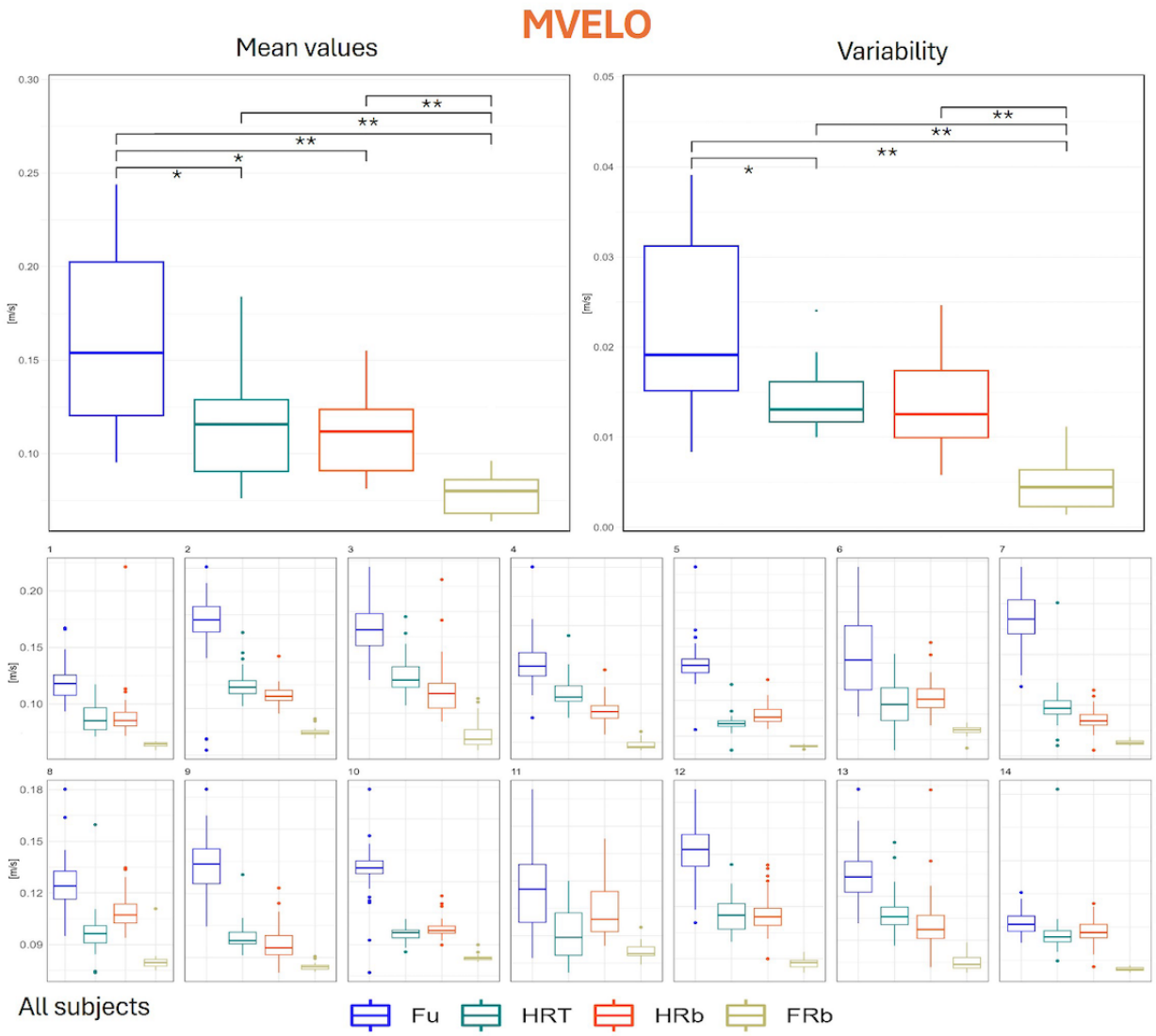
MDIST, an overall measure of excursion of the CoP from its equilibrium position, was clearly and significantly higher in Fu condition (4.21, SD 1.30 cm) than in all other conditions, and this was confirmed also at the individual level for almost every participant. In addition, as shown in Figure 6, FRb condition caused significantly lower values (1.55, SD 0.50 cm) than both HRT (2.50, SD 0.67 cm) and HRb (2.38, SD 0.59 cm), and this appeared also in all individuals but 2. No significant difference was observed between HRT and HRb conditions, as the result of different behavior across participants. Regarding intraindividual variability, we observed a reduced variability of MDIST across most participants in FRb condition (0.34, SD 0.22 cm) as compared with HRb (0.59, SD 0.23 cm), HRT (0.65, SD 0.37 cm), and Fu (0.85, SD 0.36 cm).

Figure 6. Distribution of MDIST values. Upper panel, left: group mean values across the 4 collaboration modalities; upper panel, right: distribution of intraindividual variability (SD across repetitions); lower panels: distribution of values for each participant. * denotes $P<.05$ as obtained from inferential statistics; ** denotes $P<.01$ as obtained from inferential statistics. FRb: full robot; Fu: full; HRb: half robot; HRT: half robot touch; MDIST: mean distance.



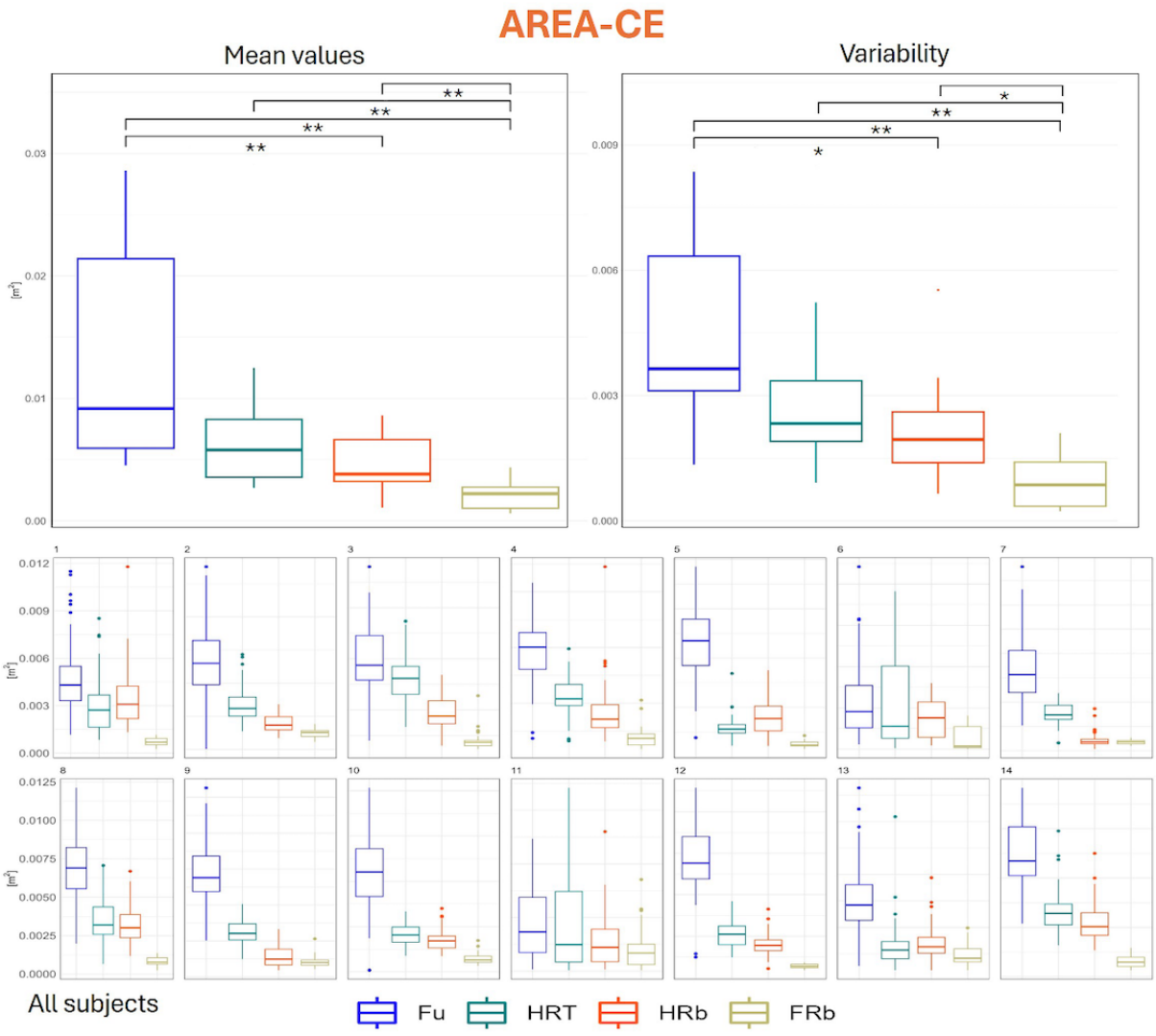
MVELO, an indicator of the amount of movement described by the CoP trajectory, in general confirms what already reported for MDIST: Fu condition determined higher values (16.3, SD 5.2 cm/s) than all other collaboration conditions, and in FRb, MVELO was significantly lower (7.9, SD 1.1 cm/s) than in all other conditions (11.1, SD 2.2 cm/s for HRb; 11.3, SD 2.9 cm/s for HRT). As depicted in Figure 7, these figures were confirmed also at the individual level for all but 2 participants. In terms of variability, while we observed that it was mostly reduced in the FRb condition (0.5, SD 0.3 cm/s) as compared with all the other conditions (2.3, SD 1.0 cm/s for Fu; 1.4, SD 0.5 cm/s for HRb), a significant decrease of the variability was present also when moving from Fu to HRT (1.4, SD 0.4 cm/s). Other variations were not significant at the group level.

Figure 7. Distribution of MVELO values. Upper panel, left: group mean values across the 4 collaboration modalities; upper panel, right: distribution of intraindividual variability (SD across repetitions); lower panels: distribution of values for each participant. * denotes *P*<.05 as obtained from inferential statistics; ** denotes *P*<.01 as obtained from inferential statistics. FRb: full robot; Fu: full; HRb: half robot; HRT: half robot touch; MVELO: mean velocity.



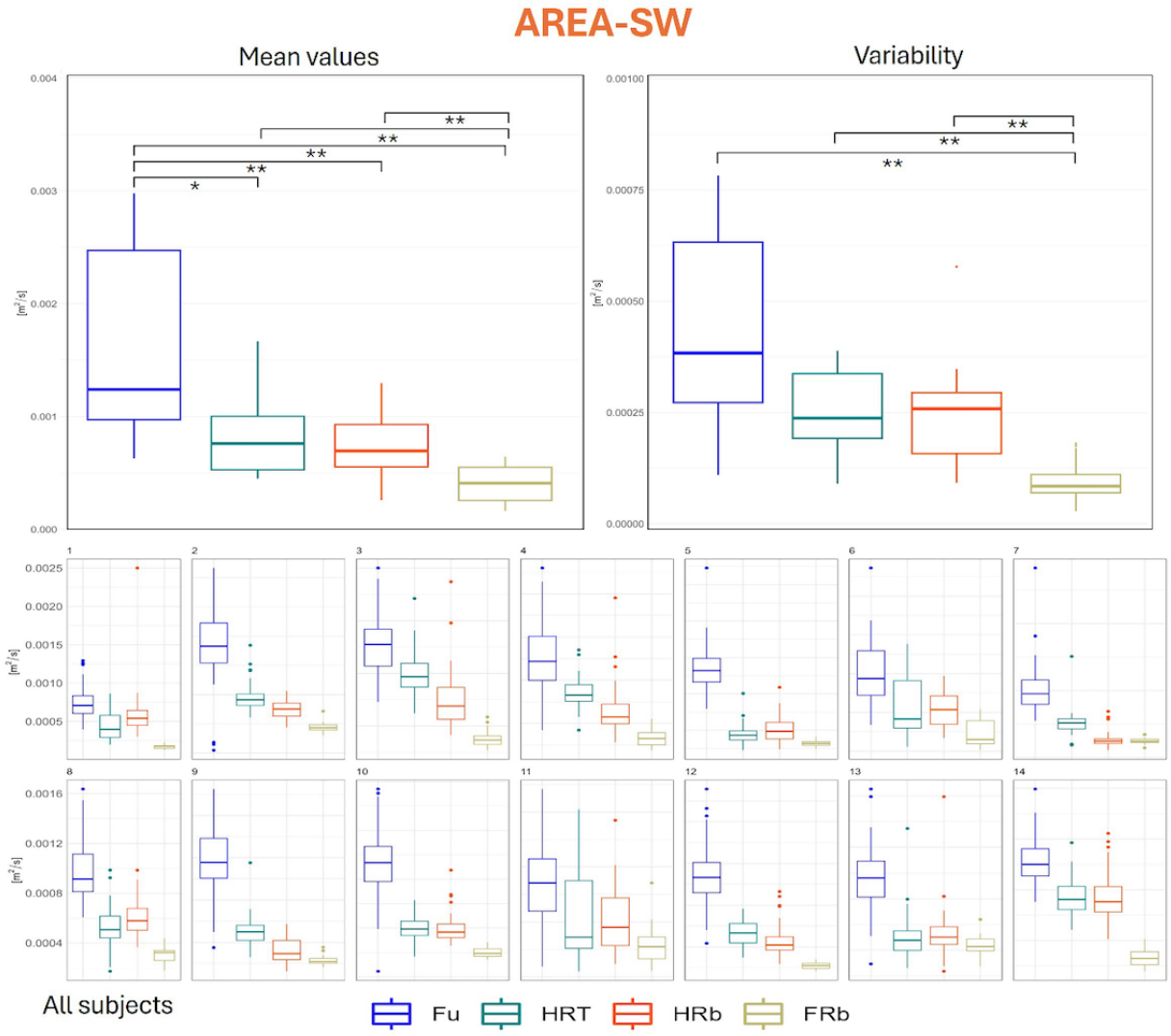
AREA-CE, which quantifies the planar extent of coverage of the CoP, displayed a behavior similar to what was seen for MVELO: values resulted significantly higher in Fu (134, SD 91 cm²), and lower in FRb (22, SD 12 cm²) than in all the other conditions (64, SD 32 cm² for HRT; 47, SD 24 cm² for HRb), respectively (Figure 8). Again, FRb intraindividual variability was lower in FRb (10, SD 7 cm²) than in all other conditions, and we saw a reduced variability when passing from Fu (45, SD 23 cm²) to HRb (21, SD 12 cm²).

Figure 8. Distribution of AREA-CE values. Upper panel, left: group mean values across the 4 collaboration modalities; upper panel, right: distribution of intraindividual variability (SD across repetitions); lower panels: distribution of values for each participant. * denotes *P*<.05 as obtained from inferential statistics; ** denotes *P*<.01 as obtained from inferential statistics. AREA-CE: confidence ellipse area; FRb: full robot; Fu: full; HRb: half robot; HRT: half robot touch.



Finally, AREA-SW, a planar measure that combines CoP extension and velocity, displayed behavior similar to what obtained with MDIST: Fu caused significantly higher AREA-SW values (16.2, SD 8.4 cm²/s) than all the other conditions (8.4, SD 3.7 cm²/s and 7.3, SD 2.8 cm²/s for HRT and HRb, respectively), and again FRb brought to values lower (4.0, SD 1.6 cm²/s) than all the other collaboration modalities (Figure 9). Both these results were reflected at the individual level for all participants but one. Intraindividual variability of AREA-SW was significantly lower in FRb (0.9, SD 0.4 cm²/s) than in all other conditions (4.4, SD 2.2 cm²/s; 2.5, SD 0.8 cm²/s; and 2.5, SD 1.2 cm²/s for Fu, HRT, and HRb, respectively).

Figure 9. Distribution of AREA-SW values. Upper panel, left: group mean values across the 4 collaboration modalities; upper panel, right: distribution of intraindividual variability (SD across repetitions); lower panels: distribution of values for each participant. * denotes *P*<.05 as obtained from inferential statistics; ** denotes *P*<.01 as obtained from inferential statistics. AREA-SW: sway area; FRb: full robot; Fu: full; HRb: half robot; HRT: half robot touch.



Discussion

Principal Findings

The analysis of the CoP trajectories shows that considering data obtained in tests going from the absence of cobot intervention (Fu) to the ones with maximum cobot assistance (FRb), the area of the CoP trajectory significantly decreases, thus possibly indicating a lower amount of oscillations of the body when the cobot intervenes.

As a general consideration, CoP trajectories, when no action of the cobot is present, tend to show greater intensity toward the loading area, indicating a higher amount of sway. This suggests that lifting affects balance control strategies. On the other hand, in both tests involving partial cooperation with the cobot, the CoP draws trajectories mainly toward the arrival station, even in the HRT modality, where the participant's arm moves toward the starting station direction to ask the cobot for help. This aspect suggests that the movement of the arm while carrying the load notably influences the CoP trajectory, while the arm movement, required to ask the cobot for help occurring in the HRT condition, does not produce a significant influence on the CoP trajectory, as compared with the modality where the cobot independently carries the load to the scanning zone (ie, HRb collaboration modality).

Regarding the detailed analysis of the specific CoP trajectories of all subjects and for all test types, results show that during tasks performed alone, CoP paths can be classified into two types: (1) concentrated with symmetrical movements toward both the starting and arrival stations (right and left), and (2) more scattered with a tendency to oscillate between the central rest zone and the arrival station (left). The variability of CoP trajectories across individuals may confirm the adoption of different independent control strategies, as also identified by the 2 types listed before. When instead maximum assistance by the cobot is present, areas drawn by CoP are smaller, and with a reduced interindividual variability. This homogenization effect can be explained by the fact that the cobot sets the activity's pace, causing the subject to adopt a more uniform behavior and timing. This is both apparent at the individual level (ie, each

subject tends to maintain a more uniform CoP trajectory across repetitions), and at the group level.

Results coming from the qualitative description of CoP trajectories were also confirmed by statistical analysis of CoP-derived parameters: specifically, CoP tends to travel more (MVELO and, to a certain extent, AREA-SW) and further (MDIST and AREA-CE) if the level of intervention from the cobot increases. In general, smaller excursions can be associated with improved balance: this means a higher tendency to remain in a smaller (and generally more stable) area during task repetitions. A possible interpretation is that while the cobot influences task execution, it also makes the operator move less with reduced postural oscillation. This aspect is relevant as an increased excursion from the equilibrium position and increased movement of the CoP are associated with an increased risk of falling [28], and incorrect postures in the working scenario are agreed to lead to the development of MSDs [29]. As seen from this perspective only, the modality of full assistance by the cobot brings to the lowest risk, as it clearly leads to minimal excursions and movements of the body.

However, we also observed a marked reduction of intraindividual variability of these parameters when full assistance from the cobot is present: one hypothesis for this evidence is that the presence of the cobot reduces autonomy in task execution and may thus impose a control strategy which is not necessarily beneficial (or chosen) by the operator. If a reduction of variability of such parameters from the baseline condition may be seen as an indicator of long-term discomfort, this needs to be appropriately considered when designing the most beneficial collaboration modalities. In this regard, intermediate collaboration modalities seem to incorporate a tradeoff between the advantages in the carrying task compensation given by the cobot on the load and the possibility of granting autonomy to the operator during the task execution.

Regarding the 2 collaboration modalities that maintain a material operator intervention, statistical analysis at the group level was rather inconclusive: in excursion-based CoP parameters, no difference was found, and this is counterintuitive, considering that one of the modalities required participants to raise the arm and kinematically reach out to ask for assistance. This was not reflected in most CoP-derived parameters, and the possible explanation for this absence of differences may be ascribed to the inherent nature of the CoP variable: it is only indirectly associated with whole body movement: coming from the ground reaction forces, it is a measure of where the foot pressure is concentrated [30]. In static conditions, CoP basically tracks the body barycenter projection to the ground (and dynamically overtakes it in specific conditions) to counteract balance disequilibrium. According to the inverted pendulum hypothesis, the body barycenter projection is thus a filtered version of the CoP trajectory. However, with self-initiated body part movements, the central nervous system is able to predict future whole-body excursions [31], and the position of the CoP may not necessarily reflect the excursions of the body barycenter. If this were the case, activation of the relevant muscles (typically, tibialis anterior) would appear in anticipation of the upper limb movement.

Regarding the inconclusive results appearing between the 2 modalities where collaboration was midlevel (ie, HRb and HRT), we tend to ascribe this process to a rather diverse behavior observed at the individual level: decreases of measures of CoP excursion and movement were in some individuals clear when decreasing the amount of user intervention (ie, passing from HRT to HRb), however this effect disappears in a nonnegligible share of individuals; as it was apparent from the qualitative description of CoP trajectories, 2 different group behaviors seemed to emerge: in one group, CoP is maintained compact despite the lesser level of cobot intervention as compared with FRb, while the second group displays higher amounts of excursion and movement. We were not able to ascribe this phenomenon to specific individual characteristics; whole body kinematics and electromyography activity analysis may help shed light on this, in particular, to check if this evidence might be associated with task role-taking, which may be a key factor in HRC activities [32].

Reduced postural variability, as evidenced by the decreased CoP trajectory area and reduced inter and intraindividual differences under maximum cobot assistance, can have significant long-term implications for musculoskeletal health. A prolonged state of limited variability in posture and movement, while potentially stabilizing in the short term, may contribute to the development of MSDs over time. This is because variability in postural strategies allows the redistribution of mechanical loads across different tissues and joints, reducing the risk of overuse injuries or chronic strain on specific musculoskeletal structures. Conversely, homogenized and repetitive postures imposed by external systems, such as cobots, may lead to localized fatigue, reduced muscular engagement, and impaired adaptability of the neuromuscular system. Consequently, these factors can heighten the susceptibility to cumulative trauma disorders, particularly in dynamic occupational scenarios. These findings highlight the importance of designing collaborative robotic systems that balance task support with opportunities for natural and varied movement patterns to mitigate potential long-term health risks for operators.

The implications of these findings for industrial settings may be interesting. The integration of cobots can significantly enhance worker safety and productivity by reducing physical strain and improving postural control. By enabling safer and more ergonomic working conditions, cobots can play a crucial role in creating healthier work environments, reducing the incidence of work-related MSDs, and improving worker satisfaction.

In general, referring to the obtained results, it must be remarked that the tasks analyzed were simulated rather than conducted in real-world contexts, so further analysis could be conducted to generalize the findings. However, the tests conducted in this study, even if simulating a cashier task, represent real work tasks themselves, and this supports the hypothesis that similar results can be obtained in the real world. Additionally, the sample consisted exclusively of young male participants, so a direct validation on female subjects and older ones was not conducted. Rather than a limit, this might be seen as a validation for the examined population, although it can be extended to other worker categories.

Conclusions

This study investigated the biomechanical impact of various levels of HRC on postural control during working tasks while standing, providing substantial insights into the benefits of cobot assistance for workers' safety and ergonomics. Through an analysis of CoP trajectories and related parameters, it was found that higher levels of cobot collaboration, particularly in the HRT, HRb, and FRb modalities, significantly improved postural stability and reduced biomechanical load on workers. These findings suggest a decreased risk of MSDs, as evidenced by lower values in CoP-derived parameters when cobot assistance increases. The study also highlighted that the intraindividual variability of CoP parameters diminished when levels of cobot involvement were increased: this indicates on one side that cobot integration promotes more uniform ergonomic benefits across different individuals, but at the same time it may tend to impose a similar behavior in balance control which can lead subjects to adopt less ecological strategies.

Moreover, while participants can set autonomously their repetition pace, HRC activities will determine an increased homogenization of repetition rhythm (or parts of it) across workers, and this may potentially lead to decrease ergonomics, as subjects can feel forced to perform the required tasks based on timings dictated by HRC. Intermediate collaboration modalities, such as HRT, can provide both advantages in terms of reduction of CoP-derived parameters—that can be indirectly associated with a reduced biomechanical risk—with the possibility of selecting a more comfortable rhythm in executing tasks.

However, further research is needed to explore the long-term effects of cobot assistance on worker health and productivity. In particular, future studies should include other sensing technologies, able to capture the neuromechanics of HRC. In addition, application to diverse populations and real-world industrial settings to generalize the results and provide a more comprehensive understanding of the benefits and potential challenges of cobot integration. Additionally, exploring the psychological and social impacts of working with cobots on employees would provide a holistic view of the implications of collaborating scenarios in industrial environments: if properly integrated into working environments cobots can potentially revolutionize industrial work by enhancing safety, efficiency, and ergonomic well-being, facilitating the development of more intelligent and flexible workrooms and working conditions.

Acknowledgments

This work was partially supported by National Institute for Insurance against Accidents at Work (INAIL): BRIC2019-BRISK and BRIC2022-CoRoMan (CUP: F83C22002780001). Activities of MS, SR, and SC were also supported by Project ECS0000024 Rome Technopole (CUP: F83B22000040006).

Authors' Contributions

DB, GC, MS, and SC contributed to the conceptualization of the study. DB developed the methodology and designed the experimental setup. GC was responsible for software development. DB, GC, and MS performed validation. Formal analysis and visualization were conducted by DB. Data curation was carried out by DB, GC, and SR. DB prepared the original draft, while DB, GC, MS, SR, and SC participated in reviewing and editing the manuscript. Supervision was provided by DB and SC. MS and SC oversaw project administration, and SC secured funding for the study.

Conflicts of Interest

None declared.

References

1. Kagermann H, Helbig J, Wahlster W. Recommendations for implementing the strategic initiative INDUSTRIE 4.0: securing the future of german manufacturing industry. : Forschungsunion; 2013 URL: https://www.din.de/resource/blob/76902/e8cac883f42bf28536e7e8165993f1fd/recommendations-for-implementing-industry-4-0-data.pdf [accessed 2025-02-17]
2. Todescato M, Braholli O, Chaltsev D, et al. Sustainable manufacturing through application of reconfigurable and intelligent systems in production processes: a system perspective. Sci Rep 2023 Dec 16;13(1):22374. [doi: 10.1038/s41598-023-49727-5] [Medline: 38104211]
3. Ajoudani A, Albrecht P, Bianchi M, et al. Smart collaborative systems for enabling flexible and ergonomic work practices [industry activities]. IEEE Robot Automat Mag 2020;27(2):169-176. [doi: 10.1109/MRA.2020.2985344]
4. Kadir BA, Broberg O. Human-centered design of work systems in the transition to industry 4.0. Appl Ergon 2021 Apr;92:103334. [doi: 10.1016/j.apergo.2020.103334] [Medline: 33264676]
5. Keshvarparast A, Battini D, Battaia O, Pirayesh A. Collaborative robots in manufacturing and assembly systems: literature review and future research agenda. J Intell Manuf 2024 Jun;35(5):2065-2118. [doi: 10.1007/s10845-023-02137-w]
6. Villani V, Pini F, Leali F, Secchi C. Survey on human–robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatronics (Oxf) 2018 Nov;55:248-266. [doi: 10.1016/j.mechatronics.2018.02.009]
7. Realyvásquez-Vargas A, Cecilia Arredondo-Soto K, Luis García-Alcaraz J, Yail Márquez-Lobato B, Cruz-García J. Introduction and configuration of a collaborative robot in an assembly task as a means to decrease occupational risks and

increase efficiency in a manufacturing company. Robot Comput Integr Manuf 2019 Jun;57:315-328. [doi: 10.1016/j.rcim.2018.12.015]
8. Cherubini A, Navarro B, Passama R, et al. Interdisciplinary evaluation of a robot physically collaborating with workers. PLoS ONE 2023;18(10):e0291410. [doi: 10.1371/journal.pone.0291410] [Medline: 37819889]
9. Johanson DL, Ahn HS, MacDonald BA, et al. The effect of robot attentional behaviors on user perceptions and behaviors in a simulated health care interaction: randomized controlled trial. J Med Internet Res 2019 Oct 4;21(10):e13667. [doi: 10.2196/13667] [Medline: 31588904]
10. Fiorini L, D'Onofrio G, Sorrentino A, et al. The role of coherent robot behavior and embodiment in emotion perception and recognition during human-robot interaction: experimental study. JMIR Hum Factors 2024 Jan 26;11:e45494. [doi: 10.2196/45494] [Medline: 38277201]
11. Yerebakan MO, Gu Y, Gross J, Hu B. Evaluation of biomechanical and mental workload during human-robot collaborative pollination task. Hum Factors 2025 Feb;67(2):100-114. [doi: 10.1177/00187208241254696] [Medline: 38807491]
12. Lorenzini M, Lagomarsino M, Fortini L, Gholami S, Ajoudani A. Ergonomic human-robot collaboration in industry: a review. Front Robot AI 2022;9:813907. [doi: 10.3389/frobt.2022.813907] [Medline: 36743294]
13. Gualtieri L, Fraboni F, De Marchi M, Rauch E. Development and evaluation of design guidelines for cognitive ergonomics in human-robot collaborative assembly systems. Appl Ergon 2022 Oct;104:103807. [doi: 10.1016/j.apergo.2022.103807] [Medline: 35763990]
14. Gualtieri L, Fraboni F, Brendel H, Pietrantoni L, Vidoni R, Dallasega P. Updating design guidelines for cognitive ergonomics in human-centred collaborative robotics applications: an expert survey. Appl Ergon 2024 May;117:104246. [doi: 10.1016/j.apergo.2024.104246] [Medline: 38354552]
15. Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O. Progress and prospects of the human–robot collaboration. Auton Robot 2018 Jun;42(5):957-975. [doi: 10.1007/s10514-017-9677-2]
16. Merlo E, Lamon E, Fusaro F, et al. An ergonomic role allocation framework for dynamic human–robot collaborative tasks. J Manuf Syst 2023 Apr;67:111-121. [doi: 10.1016/j.jmsy.2022.12.011]
17. Guideline for introducing and implementing real-time instrumental-based tools for biomechanical risk assessment. : European Committee for Standardization (CEN); 2023.
18. Ranavolo A, Ajoudani A, Cherubini A, et al. The sensor-based biomechanical risk assessment at the base of the need for revising of standards for human ergonomics. Sensors (Basel) 2020 Oct 10;20(20):5750. [doi: 10.3390/s20205750] [Medline: 33050438]
19. McDevitt S, Hernandez H, Hicks J, et al. Wearables for biomechanical performance optimization and risk assessment in industrial and sports applications. Bioengineering (Basel) 2022 Jan 13;9(1):33. [doi: 10.3390/bioengineering9010033] [Medline: 35049742]
20. Lind CM, Abtahi F, Forsman M. Wearable motion capture devices for the prevention of work-related musculoskeletal disorders in ergonomics-an overview of current applications, challenges, and future opportunities. Sensors (Basel) 2023 Apr 25;23(9):4259. [doi: 10.3390/s23094259] [Medline: 37177463]
21. Ranaldi S, Bibbo D, Corvini G, Schmid M, Conforto S. Modalities of sequential human robot collaboration trigger different modifications of trunk oscillations. Front Neurorobot 2023;17:1183164. [doi: 10.3389/fnbot.2023.1183164] [Medline: 37425334]
22. Corvini G, Ajoudani A, Conforto S, et al. Assessing biomechanical risks in human-robot collaboration: analysis of muscle activity with different intervention conditions. Gait Posture 2023 Oct;105:S18-S19. [doi: 10.1016/j.gaitpost.2023.07.308]
23. Varrecchia T, Chini G, Tarbouriech S, et al. The assistance of BAZAR robot promotes improved upper limb motor coordination in workers performing an actual use-case manual material handling. Ergonomics 2023 Dec;66(12):1950-1967. [doi: 10.1080/00140139.2023.2172213] [Medline: 36688620]
24. Winter D. Human balance and posture control during standing and walking. Gait Posture 1995 Dec;3(4):193-214. [doi: 10.1016/0966-6362(96)82849-9]
25. Draicchio F, Trebbi M, Mari S, et al. Biomechanical evaluation of supermarket cashiers before and after a redesign of the checkout counter. Ergonomics 2012;55(6):650-669. [doi: 10.1080/00140139.2012.659762] [Medline: 22455556]
26. Schmid M, Conforto S, Camomilla V, Cappozzo A, D'Alessio T. The sensitivity of posturographic parameters to acquisition settings. Med Eng Phys 2002 Nov;24(9):623-631. [doi: 10.1016/S1350-4533(02)00046-2]
27. Prieto TE, Myklebust JB, Hoffmann RG, Lovett EG, Myklebust BM. Measures of postural steadiness: differences between healthy young and elderly adults. IEEE Trans Biomed Eng 1996 Sep;43(9):956-966. [doi: 10.1109/10.532130] [Medline: 9214811]
28. Piirtola M, Era P. Force platform measurements as predictors of falls among older people - a review. Gerontology 2006;52(1):1-16. [doi: 10.1159/000089820] [Medline: 16439819]
29. Gómez-Galán M, Pérez-Alonso J, Callejón-Ferre Á, López-Martínez J. Musculoskeletal disorders: OWAS review. Ind Health 2017 Aug 8;55(4):314-337. [doi: 10.2486/indhealth.2016-0191] [Medline: 28484144]
30. Gage WH, Winter DA, Frank JS, Adkin AL. Kinematic and kinetic validity of the inverted pendulum model in quiet standing. Gait Posture 2004 Apr;19(2):124-132. [doi: 10.1016/S0966-6362(03)00037-7]

31. Kaewmanee T, Aruin AS. The role of predictability of perturbation in control of posture: a scoping review. Motor Control 2022 Jan 1;26(1):97-143. [doi: 10.1123/mc.2021-0074] [Medline: 34891127]
32. Sacheli LM, Tidoni E, Pavone EF, Aglioti SM, Candidi M. Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. Exp Brain Res 2013 May;226(4):473-486. [doi: 10.1007/s00221-013-3459-7] [Medline: 23503771]

Abbreviations

AP: anterior-posterior
AREA-CE: confidence ellipse area
AREA-SW: sway area
cobot: collaborative robot
CoP: center of pressure
FRb: full robot
Fu: full
HRb: half robot
HRC: human-robot collaboration
HRT: half robot touch
MDIST: mean distance
ML: medial-lateral
MSD: musculoskeletal disorder
MVELO: mean velocity
PPM: packages per minute

Edited by A Kushniruk; submitted 30.07.24; peer-reviewed by A Ranavolo, J Walsh; revised version received 05.12.24; accepted 20.12.24; published 27.02.25.

Please cite as:
Bibbo D, Corvini G, Schmid M, Ranaldi S, Conforto S
The Impact of Human-Robot Collaboration Levels on Postural Stability During Working Tasks Performed While Standing: Experimental Study
JMIR Hum Factors 2025;12:e64892
URL: https://humanfactors.jmir.org/2025/1/e64892
doi:10.2196/64892

User Experience of a Bespoke Videoconferencing System for Web-Based Family Visitation for Patients in an Intensive Care Unit: 1-Year Cross-Sectional Survey of Nursing Staff

Aoife Murray[1,2], MBChB, MSc; Irial Conroy[1], BEng, MSc; Frank Kirrane[3], BSc, MSc; Leonie Cullen[4], CNM; Hemendra Worlikar[1], MSc; Derek T O'Keeffe[2,5], BEng, MBChB, MSc, MD, PhD

[1]Health Innovation Via Engineering Laboratory, School of Medicine, University of Galway, University Road, Galway, Ireland

[2]School of Medicine, College of Medicine Nursing and Health Sciences, University of Galway, Galway, Ireland

[3]Department of Medical Physics and Clinical Engineering, University Hospital Galway, Galway, Ireland

[4]Critical Care Department, University Hospital Galway, Galway, Ireland

[5]Lero Science Foundation Ireland Centre for Software Research, University of Galway, Galway, Ireland

Corresponding Author:
Aoife Murray, MBChB, MSc
Health Innovation Via Engineering Laboratory, School of Medicine, University of Galway, University Road, Galway, Ireland

Abstract

Background: During the COVID-19 pandemic, in-person visitation within hospitals was restricted and sometimes eliminated to reduce the risk of transmission of SARS-CoV-2. Many health care professionals created novel strategies that were deployed to maintain a patient-centered approach. Although pandemic-related restrictions have eased, these systems, including videoconferencing or web-based bedside visits, remain relevant for visitors who cannot be present due to other reasons (lack of access to transport, socioeconomic restraints, geographical distance, etc).

Objective: The aims of this study were (1) to report the experience of intensive care nursing staff using a bespoke videoconferencing system called ICU FamilyLink; (2) to examine the scenarios in which the nursing staff used the system; and (3) to assess the future use of videoconferencing systems to enhance communication with families.

Methods: A modified Telehealth Usability questionnaire was administered to the nursing staff (N=22) of an intensive care unit in a model 4 tertiary hospital in Ireland 1 year after implementing the bespoke videoconferencing system.

Results: In total, 22 nurses working in the intensive care department at University Hospital Galway, Ireland, responded to the survey. A total of 23% (n=5) of participants were between the ages of 25 and 34 years, 54% (n=12) were between 35 and 44 years, and 23% (n=5) were between 45 and 54 years. Most (n=15, 68%) of the participants reported never using videoconferencing in the intensive care setting to communicate with family members before March 2020. The modified Telehealth Usability Questionnaire showed overall satisfaction scores for each subcategory of ease of use and learnability, interface quality, interaction quality, reliability, satisfaction and future use, and usefulness. In total, 21 (95%) participants agreed or strongly agreed with the statement, “I would use the ICU FamilyLink system in future circumstances in which family members cannot be physically present (ie, pandemics, abroad, inability to travel, etc),” and 1 participant responded neutrally. One participant highlighted a common scenario in intensive care settings in which a videoconferencing system can be used “Even without COVID, web-based communication is important when patients become unexpectedly ill and when families are abroad.”

Conclusions: This study provides valuable insights into health care professionals’ experience using a videoconferencing system to facilitate web-based visits for families. We conclude that videoconferencing systems when appropriately tailored to the environment with the users in mind can be an acceptable solution to maintain communication with family members who cannot be physically present at the bedside. The bespoke videoconferencing system had an overall positive response from 22 nursing staff who interacted with the system at varying frequency levels.

(JMIR Hum Factors 2025;12:e54560) doi:10.2196/54560

KEYWORDS

telemedicine; health; telehealth; videoconferencing; web-based; usability; intensive care; critical care; communication; COVID-19; SARS-COV-2; intensive care unit; ICU; cross-sectional survey; nursing; transmission; transmission risk; usability questionnaire; questionnaire; reliability; satisfaction; usefulness; family

Introduction

Background

In recent years, many intensive care departments have adopted a patient- and family-centered care approach. This includes having close family members physically present at the bedside during day-to-day activities and attending care meetings. Previous studies have highlighted the psychological and physical benefits of this approach for patients and their families during an individual's inpatient stay in intensive care [1-3]. However, having family members physically present at the patient's bedside can be difficult or even impossible due to geographical distance, financial challenges, and other family obligations. A significant barrier to bedside visits in the recent past was the COVID-19 pandemic.

The pandemic imposed severe and often complete restrictions on visits and forced health care professionals to change their traditional way of delivering patient care [4]. Prior to COVID-19, critical care nursing staff had been striving toward involving relatives in the intensive care unit (ICU) setting. However, many ICUs suspended these practices with the imposition of visitor restrictions [5]. These restrictive policies were introduced to decrease the risk of transmission of COVID-19 in health care settings [4]. These policies have been shown to be extremely isolating for patients, distressing for family members, and even hampering clinical care provided by health care professionals [1,6]. In particular, patients who are critically ill or vulnerable are often reliant on family for support, and at times, families are surrogate decision makers for patients. Health care professionals quickly realized the impact of visitor restrictions on family-centered care in intensive care settings. They called urgently for ways to maintain communication with family members who were no longer allowed to be physically present at the bedside [7]. At the beginning of the pandemic, hospitals quickly adopted technological solutions, including videoconferencing systems, to maintain or reopen communication channels. Previous studies have shown that adopting telemedicine solutions has numerous barriers, including the user's acceptance and experience with the technology [8-10]. Other studies have shown positive patient experiences using videoconferencing systems for web-based family support during patient rounds. However, other studies identify challenges for health care professionals, including additional workload and difficulty learning and integrating new technology into clinical practice [11,12].

Prior to the pandemic, videoconferencing in the intensive care department was limited to tele-ICU and rarely used for remote family communication. At the start of the COVID-19 pandemic, ICUs worldwide adopted various modes of maintaining communication with family members who were not allowed to be physically present at the bedside due to the risk of transmission of COVID-19. Intensive care units often serve wide geographical areas, so videoconferencing systems have the potential to improve communication with patients' families and reduce the psychological effects of intensive care admission on both patients and families. Videoconferencing systems can enable family members who cannot be physically present to be more involved during a patient's intensive care journey.

Aim of the Study

The aims of this study were (1) to report the experience of intensive care nursing staff using a bespoke videoconferencing system designed for patients and staff to communicate with patients' families remotely during hospitalization in an intensive care department of a tertiary referral hospital that had no prior video-calling system in place; (2) to examine the scenarios that nursing staff used the videoconferencing system; and (3) to assess the future use of videoconferencing systems to enhance family communication in intensive care settings using a modified version of the validated Telehealth Usability Questionnaire (mTUQ) [13].

Methods

Overview of ICU FamilyLink

A bespoke videoconferencing system called "ICU FamilyLink" was developed for the intensive care environment to allow for ad hoc web-based bedside visits available 24/7 with 1 or more close family members who may be in separate households [14]. The system was designed to be easy to use and to allow health care professionals to maintain appropriate control to maintain the security and privacy required for a hospital setting. The technology was chosen to deliver a reliable connection with high-quality audio and video. Several key requirements were previously identified during the iterative process of development [14]. This system used a 23-inch touch-screen video end point mounted on a mobile unit stand with a commercial cloud videoconferencing platform (Figure 1).

Figure 1. Video end point mounted on a mobile unit stand at an intensive care unit bedside.

A simple menu was designed for ease of use and maximum video-call security. Functions such as recording, chat, and screen sharing were disabled for security and privacy. The home screen mirrored the department's naming scheme for patient bed spaces and had the minimum number of clicks to connect. Three mobile video units were available for bedside web-based visits, and a stationary unit was available in a quiet room for medical staff to meet with families digitally.

A suite of documents, email templates, and multimedia was also created to support staff training, use, and troubleshooting. In addition, volunteer IT professionals operated a helpline for family members to ensure clinical staff were not burdened by technical queries outside their clinical training scope. A Data Protection Impact Assessment was completed, and the system was in keeping with hospital protocol and General Data Protection Regulation compliant.

The system was developed and rapidly introduced, with the first call within 2 weeks and the complete system rollout in under 3 weeks. Despite this rapid rollout, significant emphasis was placed on staff training, completion of relevant hospital risk assessments, and adherence to hospital policies. Focus on these areas was essential to ensure that staff, who were the main facilitators of the system, would find the system acceptable.

Survey Development

The validated Telehealth Usability questionnaire (TUQ) was used with minor modifications to capture relevant information to the specific use case [13]. The modified questionnaire (mTUQ) was developed with input from clinicians, engineers, a medical physicist, and a senior intensive care nurse.

The survey consisted of 40 questions in total. It included 22 questions from the TUQ, using a 7-point Likert scale, to assess the ease of use and learnability, interface quality, interaction quality, reliability, satisfaction and future use, and usefulness. The remaining questions ascertained participants' demographics and details on the respondent's use of the videoconferencing system in the ICU. Free-text comment boxes were available to gather additional qualitative data.

Recruitment

The mTUQ was distributed in paper form and via a web-based survey 1 year after the ICU FamilyLink was deployed. This allowed the system to become established in the department and minimized any bias that may have been created at the peak of the pandemic. All ICU nursing staff were eligible to participate in this study. Nursing staff were the primary users and facilitators of the ICU FamilyLink system. The system was mainly used for web-based bedside visits, rather than formal medical updates from the ICU medical team, although it was available if needed. The survey was open to any ICU staff member. Staff were recruited via emails, verbal reminders at nursing handover, and paper copies of surveys left in break rooms and main staff areas.

Data Analysis

Data were exported from the web-based survey and manually inputted from the paper surveys into a Microsoft Excel spreadsheet for analysis.

Ethical Considerations

Ethics approval for this study was obtained from the Galway Hospital Clinical Research Ethics Committee (ref C.A. 2674), and written informed consent, including web-based signatures for web-based participants was obtained from all respondents. All respondents participated voluntarily, could opt out at any point, and did not receive financial compensation. All data were anonymized.

Results

A total of 22 nurses who had been working in the intensive care unit at University Hospital Galway responded to the survey. All the respondents had used the ICU FamilyLink system at least once, although prior usage was not a requirement for participation in this study.

Characteristics of Participants

All participants were nursing staff, of these, 18% (n=4) were nurse managers. In total, 23% (n=5) of participants were between the ages of 25 and 34 years, 54% (n=12) were between 35 and 44, and 23% (n=5) were between 45 and 54 years. A total of 68% (n=15) of participants reported that they had never used videoconferencing in the intensive care setting to communicate with family members prior to March 2020 (Table 1).

During the initial introduction of the system, short training sessions were held, “superusers” were trained, and an instruction manual and video training were created. The survey showed that 36% (n=8) of participants watched a colleague before independent use and 77% (n=17) asked a colleague to teach them.

Table . Baseline characteristics of respondents.

	Respondents (N=22), n (%)
Age categories (years)	
25-34	5 (23)
35-44	12 (54)
45-55	5 (23)
Role	
Staff nurse	18 (82)
Nurse manager	4 (18)
Prior experience with any videoconferencing software in the critical care setting to communicate with patient’s family members	
Yes	7 (32)
No	15 (68)
Frequency of use (number of video calls)	
1-2	6 (27)
3-5	3 (14)
6-10	4 (18)
10-20	4 (18)
>20	5 (23)

Participants’ Use of ICU FamilyLink

The system’s usage was examined based on the frequency of video calls made (Table 1) and various scenarios (Table 2). In total, 27% (n=6) of participants used the system only 1-2 times, while 23% (n=5) of participants used the system over 20 times. Participants most commonly used the system for family time with patients while the staff were present (n=19, 86%), family time with the patient without clinical staff present (n=15, 68%), and providing daily progress updates (n=10, 45%). In total, 27% (n=6) used the system for remote family support while 1 family member was physically present at the bedside. While 9% (n=2) of respondents used the system to deliver bad news, 14% (n=3) used the system for an end-of-life scenario; 1 respondent

reported using the system for a formal care team meeting with the family. In addition, respondents were asked what purposes they felt the system was not suitable for (Table 3). A total of 45% (n=10) of respondents felt that delivering bad news and end of life were not suitable scenarios for using the videoconferencing system (Table 3). One respondent commented that, "Ideally, you would give bad news in person, but it could be used if not possible."

Table . Use of the ICU FamilyLink system.

Scenarios that the respondent has used the ICU FamilyLink system	Respondents (N=22), n (%)
Family time with the patient (while clinical staff carry out routine care)	19 (86)
Family time with the patient (without clinical staff present)	15 (68)
Daily progress updates	10 (45)
Remote family support for a designated in-person family member	6 (27)
End of life	3 (14)
Giving bad news	2 (9)
Formal care team meeting or family meeting	1 (4)
Education session for family	0 (0)

Table . Scenarios that participants felt were not suitable for the ICU FamilyLink system.

Clinical scenarios	Respondents (N=22), n (%)
Giving bad news	10 (45)
End of life	10 (45)
All are suitable	7 (32)
Formal care team meeting or family meeting	3 (14)
Education session for family	2 (9)
Daily progress updates	1 (4)
Other	1 (4)
Family time with the patient (without clinical staff present)	1 (4)
Family time with the patient (while clinical staff carry out routine care)	1 (4)
Remote family support for a designated in-person family member	0 (0)

Usability Attributes of the Videoconferencing System Using the mTUQ

The mTUQ showed an overall positive satisfaction score for all subcategories, including ease of use and learnability, interface quality, interaction quality, reliability, satisfaction and future use, and usefulness (Table 4).

Table . Responses to the modified Telehealth Usability questionnaire (N=22).

	Strongly disagree, n (%)	Disagree, n (%)	Somewhat disagree, n (%)	Neutral, n (%)	Somewhat agree, n (%)	Agree, n (%)	Strongly agree, n (%)
Ease of use and learnability							
1. It was easy to learn how to set up and use the system.	0 (0)	0 (0)	2 (9)	3 (14)	6 (27)	3 (14)	8 (36)
2. It was simple to use the system.	0 (0)	0 (0)	3 (14)	2 (9)	8 (36)	2 (9)	7 (32)
3. Compared to facilitating and managing in-person visits, the FamilyLink system was a more time-efficient way to engage with families.	0 (0)	2 (9)	1 (4)	4 (18)	7 (32)	6 (27)	2 (9)
Interface quality							
4. I am able to navigate setup, initiate, and complete calls without difficulty.	0 (0)	1 (4)	2 (9)	4 (18)	4 (18)	3 (14)	8 (36)
5. The onscreen menu for FamilyLink was intuitive to navigate.	0 (0)	1 (4)	0 (0)	2 (9)	5 (23)	6 (27)	8 (36)
6. The ICU FamilyLink system could do everything I wanted it to do.	0 (0)	0 (0)	1 (4)	2 (9)	4 (18)	5 (23)	10 (45)
Interaction quality							
7. The video quality was good and provided a clear 2-way conversation between me (and the patient, when able) and the family members.	0 (0)	1 (4)	0 (0)	0 (0)	3 (14)	4 (18)	14 (64)
8. The audio quality was good and provided clear 2-way conversation between me (and the patient, if able) and their family members.	0 (0)	0 (0)	1 (4)	1 (4)	1 (4)	5 (23)	14 (64)
9. I prefer the ICU FamilyLink monitor on a stand rather than a handheld device or smaller tablet.	0 (0)	0 (0)	0 (0)	6 (27)	1 (4)	5 (23)	10 (45)

	Strongly disagree, n (%)	Disagree, n (%)	Somewhat disagree, n (%)	Neutral, n (%)	Somewhat agree, n (%)	Agree, n (%)	Strongly agree, n (%)
10. I was able to express myself effectively using the FamilyLink system.	0 (0)	0 (0)	0 (0)	1 (4)	6 (27)	5 (23)	10 (45)
11. Compared to telephone conversations, the ICU FamilyLink was better to communicate and expressing important messages.	0 (0)	1 (4)	2 (9)	2 (9)	3 (14)	6 (27)	8 (36)
Reliability							
12. The system was reliable and consistently facilitated video calls.	0 (0)	0 (0)	1 (4)	2 (9)	4 (18)	4 (18)	11 (50)
13. I feel the FamilyLink system facilitates private and secure communication.	2 (9)	2 (9)	0 (0)	3 (14)	3 (14)	8 (36)	4 (18)
14. Whenever I made a mistake using the ICU FamilyLink system, I could quickly recover.	0 (0)	0 (0)	0 (0)	4 (18)	4 (18)	10 (45)	4 (18)
15. Family members were able to follow the emailed instructions and connect to a web-based visit call without additional technical assistance from me or other health care staff.	0 (0)	1 (4)	0 (0)	2 (9)	7 (32)	6 (27)	6 (27)
Satisfaction and future use							
16. I feel comfortable communicating with my patient's family members using the FamilyLink system.	0 (0)	0 (0)	0 (0)	1 (4)	5 (23)	8 (36)	8 (36)
17. The ICU FamilyLink system is an acceptable way to communicate while visitors are restricted.	0 (0)	0 (0)	0 (0)	0 (0)	2 (9)	9 (41)	11 (50)

	Strongly disagree, n (%)	Disagree, n (%)	Somewhat disagree, n (%)	Neutral, n (%)	Somewhat agree, n (%)	Agree, n (%)	Strongly agree, n (%)
18. I would use the ICU FamilyLink system in future circumstances in which family members cannot be physically present (ie, pandemics abroad and inability to travel).	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	6 (27)	15 (68)
19. I will use the ICU FamilyLink system in addition to in-person visits in the future.	0 (0)	0 (0)	1 (4)	2 (9)	2 (9)	7 (32)	10 (45)
20. I would recommend the ICU FamilyLink system in other health care settings.	0 (0)	0 (0)	0 (0)	1 (4)	1 (4)	4 (18)	16 (73)
Usefulness							
21. The ICU FamilyLink system provides for continuity of communication while visiting is limited.	0 (0)	0 (0)	0 (0)	0 (0)	2 (9)	5 (23)	15 (68)
22. The ICU FamilyLink system has positively impacted the overall care of patients in critical care during the visitor restrictions.	0 (0)	0 (0)	0 (0)	1 (4)	3 (14)	5 (23)	13 (59)

Ease of Use and Learnability

The majority of nursing staff responded positively to ease of use and learnability questions. Interestingly, when specifically asked to compare the time efficiency of organizing in-person visits versus web-based visits (question 3), 18% (n=4) of participants responded neutrally, 14% (n=3) provided negative responses, and 68% (n=15) gave positive responses (Table 4, questions 1-3).

Two comments were made about the process of sending the videoconferencing link to family members, indicating the need to improve the process.

The equipment itself was very easy to use, it can be a pain to log out of the computer and relog in, an app on the computer at the nurses station would be better.

[Respondent who used the system 10-20 times]

I think it would be beneficial if the email sent to the families to download the software could be sent to every computer in the unit.

[Respondent who used the system more than 20 times]

These comments indicate an area for improvement to reduce the burden of work for staff members and highlight the frustration staff commonly feel when newly introduced technology adds extra tasks. There are limitations to fully integrating technology; however, decreasing the administration time for the clinical staff could lead to better technology uptake. Despite this feedback, the 2 participants reported frequent use of the system.

Interface Quality

Nursing staff most frequently responded with “strongly agree” to the interface quality questions, question 4 (n=8; 36%), question 5 (n=8; 36%), and question 6 (n=10; 45%), respectively (Table 4).

Free text comments from 2 participants included “very simple and easy to use” and “good quality.”

Interaction Quality

A total of 64% (n=14) participants strongly agreed that the video and audio quality was good and provided clear 2-way conversation. When asked about their preference for the ICU FamilyLink monitor on a stand rather than a handheld device or smaller tablet, the majority of participants (n=16, 73%) responded in the positive, only 6 (27%) participants were neutral, and none responded in the negative. Most of the respondents (n=17, 77%) felt that compared to telephone conversations, the ICU FamilyLink was better for communicating and expressing important messages (Table 4, questions 7-11).

One participant highlighted the importance of network coverage for family members, though beyond the hospital’s control, was an important factor: “Very dependent on internet coverage for the family.” Another participant indicated that the visual aspect of the system was better than phone communication: “It helped build a connection with families that isn’t possible via phone.”

Reliability

Overall, respondents more frequently responded positively to reliability (Table 4, questions 12-15). However, 4 (18%) respondents disagreed or strongly disagreed with the statement, “I felt the ICU FamilyLink system facilitates private and secure communication.”

Satisfaction and Future Use

All participants responded positively to the statement, “The ICU FamilyLink system is an acceptable way to communicate while visitors are restricted.” In total, 21 (95%) participants agreed or strongly agreed with the statement, “I would use the ICU FamilyLink system in the future circumstances in which family members cannot be physically present (i.e., pandemics, abroad, inability to travel, etc),” and 1 participant responded neutrally (Table 4, questions 16-20).

Qualitative data also indicate health care professionals’ positive feedback and wish to use the system in the future:

Even without COVID, online communication is important when patients become unexpectedly ill, and families are abroad.

Extremely valued by family members, I have had limited exposure to its use due to working mostly night shifts.

Familylink has been a lifesaver during the pandemic; in my opinion, it is very easy to use and great for patients to be able to see their relatives onscreen while visiting is restricted. I will use it at any opportunity.

Usefulness

In total, 21 (95%) participants responded positively to questions about usefulness (Table 4, questions 21 and 22), except 1 neutral response to question 22: “The ICU FamilyLink system has positively impacted the overall care of patients in intensive care during the visitor restrictions.”

Discussion

Overview

With an overall shift toward a patient- and family-centered approach, health care professionals can harness technology to enable and optimize this approach. However, with the introduction of any technology into a new environment, particularly one as complex as health care delivery, careful design, assessment, and research should be completed to ensure acceptability, usability, and impact on all users. While health care professionals are trained to a high degree in health and medicine, their training, understanding, and use of technology vary widely.

Strengths, Limitations, and Future Directions

This study included participants who had only interacted with the ICU FamilyLink system a handful of times and also very frequent users. It also was inclusive for all age ranges and had both staff grade and senior nurses. In usability studies, it is vital to capture input from all spectrums of technical users.

One limitation of this study was that it did not interview or survey patients or their families. While this was considered, due to limited staff and stressful circumstances for patients and families, it was not included as a component of the study. There have been studies recently published examining family and patient experiences in using both phone and video for web-based visits [15-17]. Staff acceptance is critical for implementing systems like the ICU FamilyLink into regular practice. ICU staff are usually required to facilitate these calls due to the dependency needs of the ICU patients.

This study, like many similar studies, was conducted under pandemic conditions. The TUQ was modified to account for these conditions. It is encouraging that responses to questions on usefulness and future use are overall positive, with 68% (n=15) of respondents “strongly agreeing” that they would use the ICU FamilyLink system in future circumstances in which family members cannot be physically present (question 18). While there are positive indications for postpandemic use and its potential value is noted by the staff members, confirming these results outside of pandemic conditions would be important.

Conclusions

The findings of this study provide valuable insights into health care professionals’ experiences using a videoconferencing system to facilitate web-based bedside visits for family members. We conclude that when appropriately tailored to the environment and with the users in mind, videoconferencing systems can be an acceptable solution for maintaining communication with family members who cannot be physically present at the bedside. Further studies are needed to better understand usability factors for technology in order to enhance and augment communication with ICU patients, staff, and patients’ family members.

Acknowledgments

The authors would like to thank all the staff and volunteers who supported the development and deployment of the ICU FamilyLink system in Galway University Intensive Care Unit as well as the participants of this study for their valuable insight and feedback about the system. This publication has resulted from research supported in part by a grant from Science Foundation Ireland and the European Regional Development Fund (grants 13/RC/2073_P2 to AM and DTOK and 13/RC/2094_P2 to DTOK).

Conflicts of Interest

None declared.

References

1. Ashana DC, Cox CE. Providing family-centered intensive care unit care without family presence-human connection in the time of COVID-19. JAMA Netw Open 2021 Jun 1;4(6):e2113452. [doi: 10.1001/jamanetworkopen.2021.13452] [Medline: 34152421]
2. van Mol MM, Boeter TG, Verharen L, Kompanje EJ, Bakker J, Nijkamp MD. Patient- and family-centred care in the intensive care unit: a challenge in the daily practice of healthcare professionals. J Clin Nurs 2017 Oct;26(19-20):3212-3223. [doi: 10.1111/jocn.13669] [Medline: 27875001]
3. Vincent JL. Critical care—where have we been and where are we going? Crit Care 2013;17(Suppl 1):S2. [doi: 10.1186/cc11500] [Medline: 23514264]
4. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Biomed 2020 Mar 19;91(1):157-160. [doi: 10.23750/abm.v91i1.9397] [Medline: 32191675]
5. McConnell B, Moroney T. Involving relatives in ICU patient care: critical care nursing challenges. J Clin Nurs 2015 Apr;24(7-8):991-998. [doi: 10.1111/jocn.12755] [Medline: 25597494]
6. Kentish-Barnes N, Cohen-Solal Z, Morin L, Souppart V, Pochard F, Azoulay E. Lived experiences of family members of patients with severe COVID-19 who died in intensive care units in France. JAMA Netw Open 2021 Jun 1;4(6):e2113355. [doi: 10.1001/jamanetworkopen.2021.13355] [Medline: 34152418]
7. Wakam GK, Montgomery JR, Biesterveld BE, Brown CS. Not dying alone—modern compassionate care in the Covid-19 pandemic. N Engl J Med 2020 Jun 11;382(24):e88. [doi: 10.1056/NEJMp2007781] [Medline: 32289215]
8. Gajarawala SN, Pelkowski JN. Telehealth benefits and barriers. J Nurse Pract 2021 Feb;17(2):218-221. [doi: 10.1016/j.nurpra.2020.09.013] [Medline: 33106751]
9. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
10. Zhang T, Mosier J, Subbian V. Identifying barriers to and opportunities for telehealth implementation amidst the COVID-19 pandemic by using a human factors approach: a leap into the future of health care delivery? JMIR Hum Factors 2021 Apr 9;8(2):e24860. [doi: 10.2196/24860] [Medline: 33779566]
11. Østervang C, Vestergaard LV, Dieperink KB, Danbjørg DB. Patient rounds with video-consulted relatives: qualitative study on possibilities and barriers from the perspective of healthcare providers. J Med Internet Res 2019 Mar 25;21(3):e12584. [doi: 10.2196/12584] [Medline: 30907746]
12. Petersson NB, Jørgensen AL, Danbjørg DB, Dieperink KB. Video-consulted rounds with caregivers: the experience of patients with cancer. Eur J Oncol Nurs 2020 Apr 21;46:101763. [doi: 10.1016/j.ejon.2020.101763] [Medline: 32447267]
13. Parmanto B, Lewis AN Jr, Graham KM, Bertolet MH. Development of the Telehealth Usability Questionnaire (TUQ). Int J Telerehabil 2016;8(1):3-10. [doi: 10.5195/ijt.2016.6196] [Medline: 27563386]
14. Conroy I, Murray A, Kirrane F, Cullen L, Anglim P, O'Keeffe D. Key requirements of a video-call system in a critical care department as discovered during the rapid development of a solution to address COVID-19 visitor restrictions. JAMIA Open 2021 Oct;4(4):ab091. [doi: 10.1093/jamiaopen/ooab091] [Medline: 34805775]
15. Galazzi A, Binda F, Gambazza S, et al. The end of life of patients with COVID-19 in intensive care unit and the stress level on their family members: a cross-sectional study. Nurs Crit Care 2023 Jan;28(1):133-140. [doi: 10.1111/nicc.12783] [Medline: 35599352]
16. Piscitello GM, Fukushima CM, Saulitis AK, et al. Family meetings in the intensive care unit during the coronavirus disease 2019 pandemic. Am J Hosp Palliat Care 2021 Mar;38(3):305-312. [doi: 10.1177/1049909120973431] [Medline: 33207937]
17. Zante B, Erne K, Jeitziner MM. Video calls did not reduce PTSD symptoms in relatives during restricted ICU visits in the COVID-19 pandemic. Sci Rep 2022 Aug 24;12(1):14405. [doi: 10.1038/s41598-022-18616-8] [Medline: 36002566]

Abbreviations

ICU: intensive care unit
mTUQ: modified Telehealth Usability Questionnaire
TUQ: Telehealth Usability Questionnaire

Edited by P Santana-Mancilla; submitted 14.12.23; peer-reviewed by N Ahmadi, PD Reyes, R Marshall; revised version received 31.07.24; accepted 12.08.24; published 21.03.25.

Please cite as:
Murray A, Conroy I, Kirrane F, Cullen L, Worlikar H, O'Keeffe DT
User Experience of a Bespoke Videoconferencing System for Web-Based Family Visitation for Patients in an Intensive Care Unit: 1-Year Cross-Sectional Survey of Nursing Staff
JMIR Hum Factors 2025;12:e54560
URL: https://humanfactors.jmir.org/2025/1/e54560
doi:10.2196/54560

The mChoice App, an mHealth Tool for the Monitoring of Preexposure Prophylaxis Adherence and Sexual Behaviors in Young Men Who Have Sex With Men: Usability Evaluation

Fabiana Cristina Dos Santos[1], MSN, RN, PhD; Maeve Brin[1], BA; Mary R Tanner[2], MD; Carla A Galindo[2], MPH; Rebecca Schnall[1], MPH, RN, PhD

[1]School of Nursing, Columbia University, 560 W 168th St, New York, NY, United States

[2]Division of HIV Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States

Corresponding Author:
Fabiana Cristina Dos Santos, MSN, RN, PhD
School of Nursing, Columbia University, 560 W 168th St, New York, NY, United States

Abstract

Background: Mobile health (mHealth) apps provide easy and quick access for end users to monitor their health-related activities. Features such as medication reminders help end users adhere to their medication schedules and automatically record these actions, thereby helping manage their overall health. Due to insufficient mHealth tools tailored for HIV preventive care in young men who have sex with men (MSM), our study evaluated the usability of the mChoice app, a tool designed to enhance preexposure prophylaxis (PrEP) adherence and promote sexual health (eg, encouraging the use of condoms and being aware of the partner's HIV status and PrEP use).

Objective: This study aimed to apply systematic usability evaluations to test the mChoice app and to refine the visualizations to better capture and display patient-reported health information.

Methods: Usability testing involved heuristic evaluations conducted with 5 experts in informatics and user testing with 20 young MSM who were taking or were eligible to take PrEP.

Results: End users demonstrated satisfaction with the appearance of the mChoice app, reporting that the app has an intuitive interface to track PrEP adherence. However, participants highlighted areas needing improvement, including chart titles and the inclusion of "undo" and "edit" buttons to improve user control when recording PrEP use.

Conclusions: Usability evaluations involving heuristic experts and end users provided valuable insights into the mChoice app's design. Areas for improvement were identified, such as enhancing chart readability and providing additional user controls. These findings will guide iterative refinements, ensuring that future versions of the app better address the needs of its target audience and effectively support HIV prevention.

(JMIR Hum Factors 2025;12:e59780) doi:10.2196/59780

KEYWORDS

HIV prevention; data visualization; patient-reported health information; mHealth; digital health; usability; human immunodeficiency virus; preexposure prophylaxis; men who have sex with men; apps; HIV; PrEP

Introduction

The incidence of HIV, although having declined modestly overall in the United States, remains a significant public health problem for men who have sex with men (MSM) [1,2]. Notably, young MSM aged 13 to 34 years represent 58% of the estimated HIV infections in 2021, indicating a pressing need for targeted intervention in this group [2]. Preexposure prophylaxis (PrEP) is a highly effective biomedical prevention strategy to reduce HIV incidence and curb the HIV epidemic [3-10]. When taken as prescribed (daily or at least 4 times per week), PrEP reduces HIV transmission by up to 99% among MSM [7-9,11-13]. While awareness of PrEP has increased among MSM [14,15], those who are most disproportionately affected have been found to be less aware of PrEP [15], and overall, widespread uptake among those with the greatest indications for PrEP remains low [14-17].

To overcome this clinical and public health challenge, technological interventions, such as mobile health (mHealth) apps, have emerged to support public health care initiatives [18]. However, despite the potential of mHealth for delivering sexual health and HIV prevention awareness, mHealth is underutilized for supporting PrEP adherence [19,20]. In response, our study team developed mChoice, an innovative PrEP adherence monitoring app, which combines the CleverCap smart pill bottle with WiseApp, allowing end users to

self-monitor and manage their medication adherence. The CleverCap pill bottle dispenses the prescribed medication and interfaces with WiseApp for real-time tracking. The app was created by Compliance Meds Technology and has been used in prior studies among persons living with HIV [10,21-25]. It was adapted for the mChoice study. The app's key functionalities include monitoring PrEP adherence, visualizing data trends, and documenting sexual behavior, particularly among young MSM from diverse backgrounds.

A notable issue in the proliferation of mHealth apps is the development of this technology with minimal input from end users, leading to poor design, inadequate consideration of user needs, and ultimately, poor usability [26]. Poorly designed apps, lacking in usability, are prone to misuse, underutilization, and failure to achieve their intended objectives.

To ensure the best utilization of mHealth apps, it is essential to understand their usability, keeping in mind target end users (eg, young MSM), tasks (eg, PrEP adherence management and sexual health tracking), and cultural contexts (eg, language and beliefs). Adherence to PrEP is crucial for its effectiveness in preventing HIV, yet many young MSM face challenges such as inconsistent medication schedules and limited access to tailored support tools. The mChoice app was designed to address these barriers by providing intuitive features, including medication reminders, adherence tracking, and functions to support sexual health decision-making. In this study, we sought to apply systematic usability evaluations to test the mChoice app and to refine the functions to better capture and display patient-reported health information.

Methods

Ethical Considerations

The Columbia University Institutional Review Board reviewed and approved all study activities (approval number AAAT8812). Participants signed informed consent forms prior to participating in usability testing. In our dataset, each participant was assigned an identifier, and information was deidentified at the point of analysis.

Heuristic Evaluation

Sample

Usability testing of the mChoice app included heuristic evaluation conducted with experts in informatics and end-user testing conducted with young MSM. Five experts were recruited as evaluators in accordance with Nielsen's recommendation to include 3 to 5 experts in usability testing [27]. Of the 5 experts, 4 had a PhD in Nursing with expertise in human-computer interaction, interface design, and usability testing. The experience of the experts varied between 7 and 23 years, with several impactful publications in the field of informatics.

Procedure

The experts were provided with the mChoice app and asked to complete a session in the app through case scenarios that represented the main functions of the system and think-aloud methodology [28]. The mChoice app required a password, and all study data were encrypted and stored on secure HIPAA-compliant servers at Columbia University. The experts were asked to describe what they were thinking, seeing, and doing as they completed the following 10 scenarios: (1) log in to the CleverCap app; (2) complete a sexual activity log; (3) mark your PrEP dose for today as "Taken"; (4) view tomorrow's pending PrEP dose; (5) edit your sexual activity log; (6) delete your sexual activity log; (7) view information on PrEP dosing over the past month; (8) send a chat to the study team; (9) watch a video about an HIV story; and (10) find PrEP information (see Figures 1 and 2 illustrating app functions and Figure 3 demonstrating the CleverCap smart pill bottle). Following their use of the app, the experts completed an online Heuristic Evaluation Checklist to assess how the system adhered to Nielsen's 10 usability principles [27] using Qualtrics XM, an online survey software. Each question ranged from 0 (not a usability problem) to 4 (usability catastrophe). A member of the research team analyzed the experts' comments regarding usability problems to identify areas of usability concern that could be targeted for improvement. The mean usability problem severity scores were calculated for each of the 10 usability heuristics.

Figure 1. Sexual activity log. A confidential and secure feature for users to record their sexual activities, such as gender of partner, whether a condom was used, and the HIV status of their partner. Users of the on-demand/intermittent (2-1-1) preexposure prophylaxis (PrEP) regimen can use this feature to trigger subsequent PrEP dose reminders following sexual activity (ie, one pill 24 hours and one pill 48 hours after sexual activity).

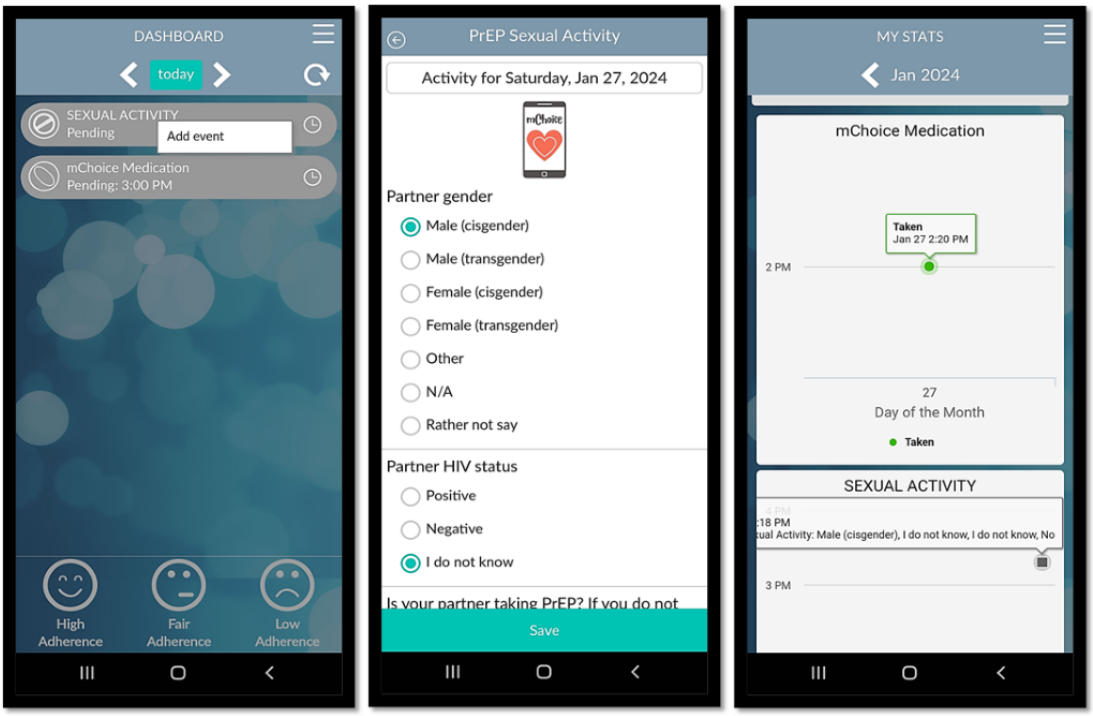


Figure 2. Preexposure prophylaxis (PrEP) adherence tracking. A simple interface for users to track PrEP intake, facilitating adherence monitoring over time. This feature is linked with the CleverCap device, a smart pill bottle that tracks when users take their medication. PrEP tracking is also customizable for users on different PrEP regimens (ie, 2-1-1, injectable, or daily oral dose) to ensure the timely intake of PrEP medication. For instance, users on injectable PrEP will be alerted of their upcoming appointment or need for their PrEP injection.

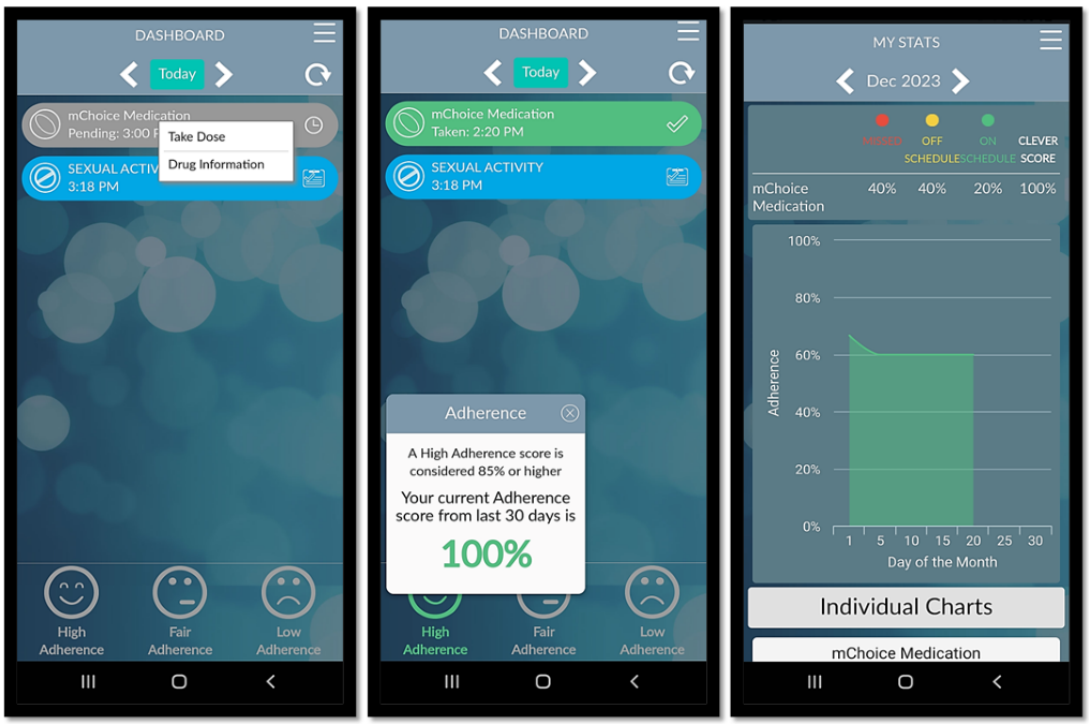


Figure 3. CleverCap device. A smart pill bottle linked with the mChoice app to track preexposure prophylaxis (PrEP) medication.

End-User Testing

Sample

Twenty end users were recruited for the usability testing. The eligibility criteria were young MSM between 18 and 39 years old, English speaking, living in New York or Alabama, and taking or eligible to take PrEP. The inclusion criteria were designed to ensure the mChoice app was evaluated by its intended target audience, ensuring a diverse usability evaluation. Potential end users were contacted using a research database and invited to participate in the study.

Procedure

During usability testing, the end users were provided with a brief explanation of the mChoice app and its functions. The end users were directed to complete the same 10 scenarios completed by the experts in the heuristic evaluation. Sessions took place remotely via Zoom meetings (Zoom Communications, Inc), and a member of the team was present during sessions to provide guidance when an end user was unable to move through a task independently and to take notes. Following their use of the app, the end users completed a survey using the Qualtrics XM software.

The survey included demographics and two validated measures of usability. The first usability measure was the self-reported ease of use measured by using the Health Information Technology Usability Evaluation Scale (Health-ITUES) [29,30]. This 20-item tool is designed to support customization at the item level to match the health information technology while retaining standardization at the construct level. It has been demonstrated to be useful for evaluating the usability of mHealth [31]. It is scored on a 5-point Likert scale ranging from 5 (strongly agree) to 1 (strongly disagree), where higher scores indicate a system that is easier to use. The second measure was the short version of the Post-Study System Usability Questionnaire (PSSUQ), a 16-item survey that assesses users' perceived satisfaction with a system [32]. Scoring is based on a 7-point Likert scale ranging from 1 (strongly agree) to 7 (strongly disagree). Lower scores indicate satisfaction with the system. Through validated measures, the end users contributed feedback on the app's functionality and user experience.

Data analysis was performed using R statistical software (version 4.1.2; R Foundation for Statistical Computing) to analyze the usability measures. The mean scores were calculated for each survey.

Results

Heuristic Evaluation

Overall, the severity scores ranged from 0.8 to 2.2 for the 10 items on the heuristic checklist, where scores closest to 0 indicate a more usable system. The mean severity scores for each heuristic item are presented in Table 1. The area identified as the most in need of improvement was "user control and freedom," where the experts identified that the app did not have an "undo" button or "edit" button for the PrEP medication entry.

The next most identified area for improvement was the "visibility of the system status," for which experts identified the lack of one "add" button to add a new sexual event. To improve the "visibility of the system status," the experts suggested we provide one "edit" button for modifying recorded sexual encounters and incorporate a separate "add" button for logging new encounters, especially in instances of multiple events. Additionally, "consistency and standards" was an area of concern with difficult terminology in charts. "Error prevention" was another concern, and experts highlighted the importance of one "edit" button to edit medication time as an issue that might impede system usability. Experts provided favorable feedback on the app's user experience. One expert (Expert 2) reported, "the app is very simple and intuitive interface…and it's nice that the button changes color with pleasant colors."

The rating score ranged from 0 (ie, best) to 4 (ie, worst), with a score of 0 indicating no usability problem, 1 indicating a cosmetic problem only, 2 indicating a minor usability problem, 3 indicating a major usability problem, and 4 indicating a usability catastrophe.

Table . Mean severity scores for items on the heuristic evaluation.

Heuristic principles	Mean (SD) severity score	Experts' comments to identify the areas of usability concerns
User control and freedom	2.2 (1.3)	• "No undo or edit for the medication entry. This is a big problem." (Expert 5)
Visibility of the system status	1.6 (1.1)	• "…When adding a sexual encounter record, it is not clear whether it edits or adds a new one, so there should be two buttons, one for edit and one to add a new one. The screen when you use edit also shows the old entry, which implies you are editing it, not starting a new one, but actually it is starting a new one. That is confusing." (Expert 3)
Consistency and standards	1.6 (1.1)	• "Not clear if the clever score is just adherence or if it means something else." (Expert 4) • "…I think in the menu it says stats. Rather than stats, it'd be adherence stats or adherence tracking over time. You know something that makes it really clear what people are clicking on." (Expert 2)
Error prevention	1.6 (1.1)	• "When entering the time of a medication, it might be easy to miss the AM/PM change. I don't think there's any design change to be made for this, but rather the user needs to be able to edit their entry if they forget to adjust this." (Expert 4)
Match between the system and the real world	1.4 (1.3)	• "Would like to see one section specific to PrEP information." (Expert 5)
Recognition rather than recall	1.4 (0.9)	• "Some of the main features can be moved to more obvious places." (Expert 1) • "Organization for videos could be improved, especially if you plan to add more content." (Expert 4)
Help and documentation	1.4 (1.3)	• "Information buttons and an introduction to the app would be helpful." (Expert 1)
Help users recognize, diagnose, and recover from errors	1.2 (1.1)	• "When the dose was missed, no way to edit if it was a mistype." (Expert 5)
Aesthetic and minimalist design	1.2 (1.1)	• "Graphs need clearer tiles and more labels." (Expert 1) • "Need more medication information or guidance if adherence is low." (Expert 5)
Flexibility and efficiency of use	0.8 (1.1)	• "…Search function for videos. Tags for topics would be good, too." (Expert 4)

End-User Testing

The end-user group for evaluating the mChoice app's usability consisted entirely of men, with the majority identifying as homosexual (14/20, 70%). The mean (SD) age was 28 (3.4) years. Half of the end users were White, and 75% (15/20) were non-Hispanic or Latino. Considering education, 50% (10/20) had a college degree. Two specific measures were employed for the usability assessment: the Health-ITUES and the PSSUQ, as detailed in Table 2.

The Health-ITUES scale, where a higher score indicates better usability, showed that "perceived ease of use" scored the highest with a mean (SD) score of 4.6 (0.5), suggesting a favorable user assessment. However, "user control" received the lowest score in this category, with a mean (SD) score of 3.8 (0.9), corroborating the heuristic evaluator's ratings.

Conversely, the PSSUQ, where a lower score indicates better usability, reflected different aspects of the mChoice performance. "System quality" scored the lowest with a mean

(SD) of 1.7 (1.2), while "interface quality" had the highest score with a mean (SD) score of 3.0 (1.8). These results suggested a positive user experience with notable ease of use and system quality strengths. However, there were areas for improvement, particularly in enhancing user control and interface quality.

Table . Usability measures.

Measures	Mean (SD) score
Health-ITUES[a]	
Perceived ease of use	4.6 (0.5)
Impact	4.1 (0.8)
Perceived usefulness	4.0 (0.8)
User control	3.8 (0.9)
Overall Health-ITUES score	4.2 (0.4)
PSSUQ[b]	
System quality	1.7 (1.2)
Information quality	2.4 (1.3)
Interface quality	3.0 (1.8)
Overall PSSUQ score	2.3 (0.6)

[a]Health-ITUES: Health Information Technology Usability Evaluation Scale (rating the score from 5 being the best score to 1 being the worst score).

[b]PSSUQ: Post-Study System Usability Questionnaire (rating the score from 1 being the best score to 7 being the worst score).

Discussion

Principal Results

Our study evaluated the mChoice app's usability in enhancing PrEP adherence for young MSM. Overall, heuristic experts and end users demonstrated satisfaction with the mChoice app, reporting that it is a simple and intuitive interface for tracking PrEP adherence. Graphs, charts, and icons are the main features of mHealth apps that help end users track their goals, habits, and achievements [33,34]. The mChoice app combines iconography (eg, emoticons) and color-coded charts to provide clear and immediate feedback on PrEP adherence. These diverse visualizations allow end users to interpret their data in several formats. The app includes emoticons in different colors (green for high adherence, red for missed doses, and yellow for off-schedule doses) to represent adherence status. For instance, a sad emoticon in red for low adherence might highlight medication adherence challenges and prompt self-reflection on areas needing improvement. Such immediate visual feedback can motivate end users to improve their medication routine. In addition, graphs showed the trend of PrEP medication adherence over time, including evidence-based information on the percentage of missed, off-schedule, and on-time doses.

While prior research has indicated that varied data visualization options help match different end users' preferences [33] and color-coded visual cues can aid in quick decision-making and interpretation of health data [35-38], our study did not directly assess end-user interpretations of these visual elements. The mChoice app incorporates these evidence-based design principles, but further research would be needed to evaluate their impact on user comprehension and medication adherence behaviors.

Our findings suggest that the visual elements available in the mChoice app offer a positive end-user experience and function accessibility that could assist in timely medication intake and monitor adherence. However, our heuristic evaluation highlighted areas needing improvement, particularly in the need for clearer chart titles and more labels to exemplify each component (eg, Clever score), underscoring the importance of refinement of the mChoice design for optimized end-user experiences.

The accuracy of the information collected through the mChoice app was another area of attention that heuristic experts raised concerns, particularly highlighting the absence of "undo," "edit," and "add" buttons. These limitations restricted user control, potentially affecting the accuracy of patients' health information input. For instance, the absence of "undo" and "edit" buttons could lead to inaccuracies if end users mistakenly enter the time of their PrEP dose intake. Additionally, the lack of an "add" button for recording multiple sexual events was seen as an important barrier, restricting end users from documenting their sexual event entries accurately. These findings underscore the need for enhanced app capabilities to ensure accurate and comprehensive user input and customizations. Similarly, other studies [39,40] reported concerns about user experiences and data accuracy in mobile apps, reinforcing the findings of our study.

The actionable finding from this study is to reevaluate the functions of the mChoice app after integrating end-user feedback and addressing heuristic violations. Iterative usability evaluations should be considered as a best practice in the app development process to ensure the effectiveness of the app's future iterations.

Limitations

A limitation of our study was the convenience sample, recruited from a database that included participants who had participated in past research and expressed interest in future studies. Additionally, while our sample targeted young MSM, English-speaking men, and those living in New York or Alabama, this demographic representation might limit our ability to understand how the app might perform for medication adherence across diverse populations. Given that PrEP users come from varied cultural, linguistic, and socioeconomic backgrounds, our findings may not be generalizable to groups whose cultural beliefs, language preferences, and health care engagement patterns differ from those of our study population. While end users were central for evaluating the usability of the mChoice app through validated quantitative measures, qualitative data were collected only from experts during heuristic evaluations. In this study, we did not evaluate the artificial intelligence or algorithmic components that could enhance personalization and medication adherence predictions. Despite these limitations, we identified areas of improvement to refine the mChoice app. In addition, this study enriched the body of research on evaluating the mHealth usability for HIV prevention. This was achieved by employing end users and heuristic experts to assess this innovative app designed for PrEP adherence and HIV prevention.

Conclusions

Our usability study of the mChoice app involved rigorous evaluations through interactive heuristic evaluations and end-user testing. The results demonstrated the app's simplicity and user-friendly interface, showcasing its potential in monitoring health-related activities such as PrEP adherence. Further research could explore how these usability enhancements might influence user engagement and behavior change to support HIV prevention efforts.

Acknowledgments

Research reported in this study was funded by the Centers for Disease Control and Prevention through grant #U01PS005229. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Conflicts of Interest

None declared.

References

1. HIV & AIDS trends and U.S. statistics overview. HIV.gov. URL: https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics [accessed 2024-02-01]
2. Estimated HIV incidence and prevalence in the United States, 2017–2021. Centers for Disease Control and Prevention. 2023. URL: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html [accessed 2023-11-12]
3. Carnegie NB, Goodreau SM, Liu A, et al. Targeting pre-exposure prophylaxis among men who have sex with men in the United States and Peru: partnership types, contact rates, and sexual role. J Acquir Immune Defic Syndr 2015 May 1;69(1):119-125. [doi: 10.1097/QAI.0000000000000555] [Medline: 25942463]
4. Juusola JL, Brandeau ML, Owens DK, Bendavid E. The cost-effectiveness of preexposure prophylaxis for HIV prevention in the United States in men who have sex with men. Ann Intern Med 2012 Apr 17;156(8):541-550. [doi: 10.7326/0003-4819-156-8-201204170-00001] [Medline: 22508731]
5. Kessler J, Myers JE, Nucifora KA, et al. Evaluating the impact of prioritization of antiretroviral pre-exposure prophylaxis in New York. AIDS 2014 Nov 28;28(18):2683-2691. [doi: 10.1097/QAD.0000000000000460] [Medline: 25493594]
6. Zablotska IB. Likely impact of pre-exposure prophylaxis on HIV epidemics among men who have sex with men. Sex Health 2017 Feb;14(1):97-105. [doi: 10.1071/SH16153] [Medline: 27883309]
7. McCormack S, Dunn DT, Desai M, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. Lancet 2016 Jan 2;387(10013):53-60. [doi: 10.1016/S0140-6736(15)00056-2] [Medline: 26364263]
8. Grant RM, Anderson PL, McMahan V, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. Lancet Infect Dis 2014 Sep;14(9):820-829. [doi: 10.1016/S1473-3099(14)70847-3] [Medline: 25065857]
9. Volk JE, Marcus JL, Phengrasamy T, et al. No new HIV infections with increasing use of HIV preexposure prophylaxis in a clinical practice setting. Clin Infect Dis 2015 Nov 15;61(10):1601-1603. [doi: 10.1093/cid/civ778] [Medline: 26334052]
10. Dos Santos FC, Batey DS, Kay ES, et al. The effect of a combined mHealth and community health worker intervention on HIV self-management. J Am Med Inform Assoc 2025 Mar 1;32(3):510-517. [doi: 10.1093/jamia/ocae322] [Medline: 39798152]
11. Effectiveness of prevention strategies to reduce the risk of acquiring or transmitting HIV. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/hiv/risk/estimates/preventionstrategies.html#anchor_1562942347 [accessed 2024-05-03]

12. Liu AY, Cohen SE, Vittinghoff E, et al. Preexposure prophylaxis for HIV infection integrated with municipal- and community-based sexual health services. JAMA Intern Med 2016 Jan;176(1):75-84. [doi: 10.1001/jamainternmed.2015.4683] [Medline: 26571482]
13. Marcus JL, Hurley LB, Nguyen DP, Silverberg MJ, Volk JE. Redefining human immunodeficiency virus (HIV) preexposure prophylaxis failures. Clin Infect Dis 2017 Oct 30;65(10):1768-1769. [doi: 10.1093/cid/cix593] [Medline: 29020235]
14. Grov C, Rendina HJ, Whitfield THF, Ventuneac A, Parsons JT. Changes in familiarity with and willingness to take preexposure prophylaxis in a longitudinal study of highly sexually active gay and bisexual men. LGBT Health 2016 Aug;3(4):252-257. [doi: 10.1089/lgbt.2015.0123] [Medline: 27183232]
15. Mayer KH, Agwu A, Malebranche D. Barriers to the wider use of pre-exposure prophylaxis in the United States: a narrative review. Adv Ther 2020 May;37(5):1778-1811. [doi: 10.1007/s12325-020-01295-0] [Medline: 32232664]
16. Dawit R, Predmore Z, Raifman J, et al. Identifying HIV PrEP attributes to increase PrEP use among different groups of gay, bisexual, and other men who have sex with men: a latent class analysis of a discrete choice experiment. AIDS Behav 2024 Jan;28(1):125-134. [doi: 10.1007/s10461-023-04131-y] [Medline: 37474623]
17. Dubov A, Ogunbajo A, Altice FL, Fraenkel L. Optimizing access to PrEP based on MSM preferences: results of a discrete choice experiment. AIDS Care 2019 May;31(5):545-553. [doi: 10.1080/09540121.2018.1557590] [Medline: 30554519]
18. Global strategy on digital health 2020 - 2025. World Health Organization. 2021. URL: https://iris.who.int/handle/10665/344249 [accessed 2024-02-01]
19. Liu AY, Laborde ND, Coleman K, et al. DOT diary: developing a novel mobile app using artificial intelligence and an electronic sexual diary to measure and support PrEP adherence among young men who have sex with men. AIDS Behav 2021 Apr;25(4):1001-1012. [doi: 10.1007/s10461-020-03054-2] [Medline: 33044687]
20. Whiteley L, Mena L, Craker LK, Healy MG, Brown LK. Creating a theoretically grounded gaming app to increase adherence to pre-exposure prophylaxis: lessons from the development of the viral combat mobile phone game. JMIR Serious Games 2019 Mar 27;7(1):e11861. [doi: 10.2196/11861] [Medline: 30916652]
21. Schnall R, Rojas M, Bakken S, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
22. Aliabadi N, Carballo-Dieguez A, Bakken S, et al. Using the information-motivation-behavioral skills model to guide the development of an HIV prevention smartphone application for high-risk MSM. AIDS Educ Prev 2015 Dec;27(6):522-537. [doi: 10.1521/aeap.2015.27.6.522] [Medline: 26595265]
23. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134. [doi: 10.2196/jmir.3393] [Medline: 24862459]
24. Schnall R, Kuhns LM, Pearson C, et al. Efficacy of MyPEEPS mobile, an HIV prevention intervention using mobile technology, on reducing sexual risk among same-sex attracted adolescent males: a randomized clinical trial. JAMA Netw Open 2022 Sep 1;5(9):e2231853. [doi: 10.1001/jamanetworkopen.2022.31853] [Medline: 36129712]
25. Beauchemin M, Gradilla M, Baik D, Cho H, Schnall R. A multi-step usability evaluation of a self-management app to support medication adherence in persons living with HIV. Int J Med Inform 2019 Feb;122:37-44. [doi: 10.1016/j.ijmedinf.2018.11.012] [Medline: 30623782]
26. McCurdie T, Taneva S, Casselman M, et al. mHealth consumer apps: the case for user-centered design. Biomed Instrum Technol 2012;Suppl:49-56. [doi: 10.2345/0899-8205-46.s2.49] [Medline: 23039777]
27. Nielsen J, Molich R. Heuristic evaluation of user interfaces. Presented at: CHI90: Conference on Human Factors in Computing; Apr 1-5, 1990; Seattle, Washington, United States p. 249-256 URL: https://dl.acm.org/doi/10.1145/97243.97281 [accessed 2025-02-19] [doi: 10.1145/97243.97281]
28. Bertini E, Gabrielli S, Kimani S, Catarci T, Santucci G. Appropriating and assessing heuristics for mobile computing. Presented at: AVI06: The International Conference on Advanced Visual Interfaces; May 23-26, 2006; Venezia, Italy URL: https://dl.acm.org/doi/10.1145/1133265.1133291 [accessed 2025-02-19] [doi: 10.1145/1133265.1133291]
29. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
30. Schnall R, Cho H, Liu J. Health Information Technology Usability Evaluation Scale (Health-ITUES) for usability assessment of mobile health technology: validation study. JMIR Mhealth Uhealth 2018 Jan 5;6(1):e4. [doi: 10.2196/mhealth.8851] [Medline: 29305343]
31. Brown W 3rd, Yen PY, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087. [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
32. Lewis JR. Psychometric evaluation of the PSSUQ using data from five years of usability studies. Int J Hum Comput Interact 2002 Sep;14(3-4):463-488. [doi: 10.1080/10447318.2002.9669130]
33. Caldeira C, Chen Y, Chan L, Pham V, Chen Y, Zheng K. Mobile apps for mood tracking: an analysis of features and user reviews. AMIA Annu Symp Proc 2017;2017:495-504. [Medline: 29854114]

34. Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. J Biomed Inform 2014 Oct;51:287-298. [doi: 10.1016/j.jbi.2014.04.006] [Medline: 24747356]
35. Albers EAC, Fraterman I, Walraven I, et al. Visualization formats of patient-reported outcome measures in clinical practice: a systematic review about preferences and interpretation accuracy. J Patient Rep Outcomes 2022 Mar 3;6(1):18. [doi: 10.1186/s41687-022-00424-3] [Medline: 35239055]
36. Kuijpers W, Giesinger JM, Zabernigg A, et al. Patients' and health professionals' understanding of and preferences for graphical presentation styles for individual-level EORTC QLQ-C30 scores. Qual Life Res 2016 Mar;25(3):595-604. [doi: 10.1007/s11136-015-1107-3] [Medline: 26353905]
37. Smith KC, Brundage MD, Tolbert E, et al. Engaging stakeholders to improve presentation of patient-reported outcomes data in clinical practice. Support Care Cancer 2016 Oct;24(10):4149-4157. [doi: 10.1007/s00520-016-3240-0] [Medline: 27165054]
38. Dos Santos FC, Yao Y, Macieira TGR, Dunn Lopez K, Keenan GM. Nurses' preferences for the format of care planning clinical decision support coded with standardized nursing languages. J Am Med Inform Assoc 2023 Oct 19;30(11):1846-1851. [doi: 10.1093/jamia/ocad093] [Medline: 37257882]
39. Gu J, Mackin S, Zheng Y. Making sense: an innovative data visualization application utilized via mobile platform. Presented at: 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS); Jun 28-30, 2018; Exeter, United Kingdom p. 1105-1109. [doi: 10.1109/HPCC/SmartCity/DSS.2018.00184]
40. Katz D, Dalton N, Holland S, O'Kane A, Price BA. Questioning the reflection paradigm for diabetes mobile apps. In: Giokas K, Bokor L, Hopfgartner F, editors. eHealth 360°: Springer, Cham; 2017:315-326. [doi: 10.1007/978-3-319-49655-9_39]

Abbreviations

Health-ITUES: Health Information Technology Usability Evaluation Scale
mHealth: mobile health
MSM: men who have sex with men
PrEP: preexposure prophylaxis
PSSUQ: Post-Study System Usability Questionnaire

Edited by A Kushniruk; submitted 16.05.24; peer-reviewed by C Grov, D Chrimes, R Ma; revised version received 19.01.25; accepted 19.01.25; published 28.02.25.

Please cite as:
Dos Santos FC, Brin M, Tanner MR, Galindo CA, Schnall R
The mChoice App, an mHealth Tool for the Monitoring of Preexposure Prophylaxis Adherence and Sexual Behaviors in Young Men Who Have Sex With Men: Usability Evaluation
JMIR Hum Factors 2025;12:e59780
URL: https://humanfactors.jmir.org/2025/1/e59780
doi:10.2196/59780

Original Paper

Use of Wearable Transdermal Alcohol Sensors for Monitoring Alcohol Consumption After Detoxification With Contingency Management: Pilot Randomized Feasibility Trial

Eileen Brobbin[1], PhD; Colin Drummond[1], MD; Stephen Parkin[2], PhD; Paolo Deluca[1], PhD

[1]King's College London, London, United Kingdom

[2]Department of Public Health, Environments and Society at London School of Hygiene and Tropical Medicine, King's College London, London, United Kingdom

Corresponding Author:
Eileen Brobbin, PhD
King's College London
4 Windswor Walk
Denmark Hill
London, SE5 8BB
United Kingdom
Phone: 44 07758287265
Email: eileen.brobbin@kcl.ac.uk

Abstract

Background: Wearable transdermal alcohol sensor (TAS) devices generate continuous data on alcohol consumption through the indiscernible sweat vapors on the skin. This continuous alcohol monitoring capability could provide a new method for alcohol services to monitor service users at various stages of their alcohol treatment.

Objective: We aimed to assess the feasibility of using a TAS as part of alcohol treatment with alcohol service users using the device with or without contingency management (CM).

Methods: A feasibility study was conducted of a convenience sample of 29 current service users from 3 South London alcohol services. Participants were randomized into either a control (treatment as usual) or CM group (treatment as usual+CM). We assessed the feasibility of enrollment, participation, device tampering and return, and device wearability and the accuracy of data capture. These data were reported descriptively where appropriate, the groups were compared, and alcohol self-report data were compared to the transdermal alcohol concentration to assess accuracy.

Results: A total of 34 individuals were approached, and 32 (94%) were enrolled and randomized (n=17, 53% to the control group and n=15, 47% to the CM group) over 5 months. In total, 3 participants withdrew (n=2, 67% from the control group and n=1, 33% from the CM group). There was a total of 203 meetings arranged (29 participants × 7 meetings), and 185 (91.1%) were attended. Only 1 of the 29 participants (3%) admitted to turning the TAS off to avoid monitoring. There were some issues with the TAS not functioning properly and not being able to be cleaned. Removals were recorded, but the definition of TAS removal may need to be improved for future trials. There was a high TAS return rate (28/29, 97% of the participants returned the TAS). Secondary outcomes suggest that the BACtrack Skyn remains an accurate tool to monitor alcohol consumption compared to self-report data and that it is acceptable to wearers over 2 weeks, with many participants (27/28, 96%) answering that they would wear it again and for longer but that the CM procedure could be made clearer.

Conclusions: The delivery of CM via a TAS was feasible in this study, but recommendations for a future larger trial include that the study design should be changed to provide an operationalized rather than manual method of checking whether TAS data meet CM criteria. This would reduce researcher burden and researcher and participant time. Current recruitment and research meeting design seem suitable for a future larger trial.

Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN) ISRCTN46845361; https://www.isrctn.com/ISRCTN46845361

(JMIR Hum Factors 2025;12:e64664) doi:10.2196/64664

KEYWORDS

alcohol; alcohol dependence; alcohol monitoring; alcohol treatment; contingency management; transdermal alcohol sensor; transdermal technology; wearable; wearable alcohol biosensor

Introduction

Background

Various wearable transdermal alcohol sensor (TAS) devices have been developed. These devices measure alcohol consumption from alcohol vapors in the skin via sweat, known as transdermal alcohol concentration (TAC), and can record at automated regular intervals. One potential use of TASs is as a tool for objective alcohol measurement in clinical alcohol treatment.

TASs could be used in specialist community alcohol services for monitoring alcohol consumption before detoxification to determine drinking levels and patterns, during detoxification, or after detoxification to be combined with sessions with key workers and other psychosocial or medication treatment. The recorded alcohol data could be used to consider triggers and provide evidence of abstinence or alcohol reduction and proof of engagement, which may be useful as evidence for funded treatment, or TASs could be used to implement contingency management (CM) for alcohol reduction. However, before TASs can be implemented into treatment, their accuracy, acceptability, and feasibility in this setting must be validated.

There have been several studies that have investigated the various TAS device brands on TAC data accuracy compared to self-report, blood alcohol concentration, and breath alcohol content data [1-17]. Most of this previous research was conducted with healthy adult volunteers [18,19]. Various brands of TASs have been developed and are in differing stages of validation and research. There have only been a few studies conducted with a specific focus on exploring TAS acceptability and feasibility measures [12,20-26] and even fewer specifically on TAS implementation of CM rewards [1,2,26-31]. While most of the TAS research reports on the SCRAM monitor, there is a growing number of individuals using a newer generation of TASs, the BACtrack Skyn [7,12-14,32-42].

CM has been evidenced as an effective treatment for substance use. It is an established treatment recommended by the National Institute for Health and Care Excellence [43] and has proven to be effective for a range of substance use treatments [44-47]. Although initially developed to be used with alcohol use disorder (AUD) treatment, it has had limited use in this area [48,49]. This is due to the nature of alcohol metabolism and its short detectability within the body [50,51].

Currently used methods for alcohol measurement include breath, blood, and urine tests, which have a relatively short time frame to detect alcohol [52]. Thus, to accurately implement CM in alcohol treatment, the individual would require frequent, multiple breath, blood, or urine tests daily to prove abstinence or alcohol reduction and accurately achieve CM rewards [47,49,53,54]. This is not always feasible with staff time and resources and increases the burden on both service users and staff. The portable mobile phone–linked breathalyzer has also tried to address this limitation and appears to work well in tracking alcohol treatment progress [55-58]. However, it still requires repeated daily compliance from the individual completing each breathalyzer test.

The development of TASs has the potential to address CM implementation with alcohol use and can provide a low-burden, low-intensity solution [1,2,26-31]. Previous literature has started to explore how TASs can implement CM [1,2,26-31]. These studies have found the SCRAM monitor successful in implementing the CM procedure and found that the CM intervention was able to significantly reduce alcohol consumption [1,2,27,29]. Of these 8 studies, none recruited alcohol-dependent participants. In total, 2 involved recent driving while intoxicated offenders with differing criteria on alcohol consumption, 1 with an Alcohol Use Disorders Identification Test score of ≥4 [28] and the other with an Alcohol Use Disorders Identification Test score of ≥8 [30]; 2 recruited HIV-diagnosed individuals with higher levels of alcohol consumption [26,31]; and the other 4 classified participants as risky or heavy drinkers with varying ways to define this [1,2,27,29]. The length of TAS wear periods and CM length ranged from 1 to 4 months, and all studies were conducted in the United States and used the SCRAM monitor.

This study followed a previous study that used the Skyn with individuals in alcohol treatment wearing the device for 1 week and interviewed staff on their thoughts [36,59,60]. From this previous study, we found that most of the individuals were willing to wear the device for longer and that staff want patients wearing the device for longer than a week for it to be used in alcohol treatment. Therefore, for this study, we extended the wear time for 2 weeks and added the CM component. The findings of this study will contribute to the knowledge on TAS feasibility in alcohol research and providing CM in alcohol services in South London.

Objectives

This study aimed to explore the feasibility, strengths, and limitations of using a TAS to monitor alcohol consumption in individuals in treatment for AUD with or without CM to promote abstinence or low-level alcohol consumption.

Methods

This was a randomized controlled trial with a 1:1 allocation ratio to the control and CM groups.

Patient and Public Involvement and Staff Consultation

To ensure that the recruitment, study design, and outcomes were appropriate, when designing this study, we conducted monthly patient and public involvement groups and staff consultation for 6 months. Service users and staff from 3 South London alcohol services were invited to attend. There were 2 meetings held each month, one for service users (in person) and one for

staff (on Microsoft Teams). In these groups, we discussed study aspects and the participants' thoughts on potential challenges.

Setting

In total, 3 drug and alcohol services were recruited from South London and Maudsley National Health Service Foundation Trust. The 3 services were the Pier Road Project (Erith), Wandsworth Community Drug and Alcohol Service (Wandsworth and Richmond), and the Assertive Alcohol Outreach Team (Camberwell). All services were willing to be involved in this study and had recruitment occur at their facilities. At each service, after referral, patients typically completed a community detoxification, were prescribed any medication if appropriate, and started to attend group meetings and regular one-to-one meetings with their key worker. All participants that were recruited for this study were currently attending groups and one-to-one sessions.

Participants and Sample Size

Participants were referred by service staff and by the researcher attending patient groups to discuss the study. One of the aims of this trial was to investigate the feasibility of conducting a larger trial. Therefore, this study conducted no formal statistical sample size calculation but aimed to recruit 30 participants (15 in each arm).

The inclusion criteria were (1) reception of alcohol treatment for an AUD in one of the participating South London alcohol services, (2) age of ≥18 years, (3) ability to speak English competently, (4) ability to meet throughout the study period, (5) no current participation in any other research trials, and (6) willingness to provide informed consent to participate. Exclusion criteria were (1) current use of any illegal or addictive substances (excluding cannabis), (2) age of <18 years, and (3) inability to speak English without a translator.

The study inclusion criteria were intentionally kept broad, enabling individuals receiving any treatment for an AUD (provided that they met the other criteria) to participate.

Randomization

Participants were randomized via dedicated software and sealed envelopes by an independent statistician. The team member recruiting participants (EB) was not aware of the randomization allocation sequence until the sealed numbered envelope was opened at each participant's first research meeting. EB enrolled all participants and assigned participants to their allocated group. The researcher and participants were not blind to the allocation; both were aware of whether they were receiving CM rewards.

Procedure

At enrollment, participants were randomized into one of two groups: (1) treatment as usual+wearing a TAS (control group) or (2) treatment as usual+wearing a TAS+CM for low or no alcohol consumption as measured using the TAS (CM group).

Each participant had 7 research meetings arranged across 15 days. At the first meeting, the participant was trained in using the TAS and provided with a quick leaflet guide to take home with them, had the study protocol explained, and was randomized into a group. Meetings 2 to 6 were for TAS data download and timeline followback (TLFB) completion. At the final meeting, meeting 7, participants also had TAS data downloaded and completed the TLFB as well as a postwear survey. This procedure is shown in Table 1.

Table 1. Study procedure—an example of the research meetings if the first meeting happened on a Monday.

	Meeting description
Week 1	
Monday	Meeting 1—talk participants through the study and train them in using the TAS[a]
Wednesday	Meeting 2—TAS data download and TLFB[b] (+CM[c])
Friday	Meeting 3—TAS data download and TLFB (+CM)
Week 2	
Monday	Meeting 4—TAS data download and TLFB (+CM)
Wednesday	Meeting 5—TAS data download and TLFB (+CM)
Friday	Meeting 6—TAS data download and TLFB (+CM)
Week 3	
Monday	Meeting 7—TAS data download, TLFB (+CM), and postwear survey

[a]TAS: transdermal alcohol sensor.

[b]TLFB: timeline followback.

[c]CM: contingency management.

If the participant was randomized into the CM group, in addition to the aforementioned steps, at each meeting (meetings 2-7), the TAS data were checked to confirm whether the participant had been abstinent or consumed an amount of alcohol below our set threshold and whether they had been wearing the TAS. If they met these 2 criteria, they were then rewarded with the corresponding amount for the days of abstinence or low level of drinking.

The reason for the research meeting design was for regular TAS data download to ensure that the TAS did not start overwriting data. At the time of this study, the Skyn could store up to 72 hours before overwriting data. Participants also completed a

TLFB with the researcher at every meeting, so they only had to recall alcohol consumption for the previous 2 to 3 days [61].

CM Intervention

Participants who were randomized to the CM group could also earn rewards by being abstinent or for low drinking as measured using the TAS and wearing the TAS consistently (we specified that removal of no more than 60 minutes per day was allowed for a shower or bath). The CM reward was a £5 (US $6.25) voucher for each day that the target behavior occurred. There were also bonuses for consecutive days of the target behavior occurring. At each meeting (meetings 2-7), the researcher checked their TAC data and provided any earned CM rewards since the previous meeting. If the participant met the target behavior every single day for the study period, there was an additional bonus given at the end (£35 [US $43.75]). Therefore, a total of £180 (US $225.02) over the study period could be given in CM vouchers. If there was a day in which this behavior did not occur, then participants received no CM for that day and were not eligible for that bonus. The CM plan is shown in Table 2.

Table 2. Contingency management (CM) plan—an example of the CM plan for a Monday start. Participants were ineligible for CM if the device was removed (skin temperature of <30 °C for >1 hour). A 1-time 1-hour removal was allowed per day.

Day number	Day	CM for 1-day abstinence (£5 [US $6.25] per day)[a]	CM bonus for consecutive-day abstinence (£5 [US $6.25] per day)[b]	CM bonus for 14 consecutive days of abstinence (£35 [US $43.75] for 14 days)[c]
1	Monday	£5 (US $6.25)	—[d]	—
2	Tuesday	£5 (US $6.25)	—	—
3	Wednesday	£5 (US $6.25)	£10 (US $12.50; second meeting)	—
4	Thursday	£5 (US $6.25)	—	—
5	Friday	£5 (US $6.25)	£10 (US $12.50; third meeting)	—
6	Saturday	£5 (US $6.25)	—	—
7	Sunday	£5 (US $6.25)	—	—
8	Monday	£5 (US $6.25)	£15 (US $18.75; fourth meeting)	—
9	Tuesday	£5 (US $6.25)	—	—
10	Wednesday	£5 (US $6.25)	£10 (US $12.50; fifth meeting)	—
11	Thursday	£5 (US $6.25)	—	—
12	Friday	£5 (US $6.25)	£10 (US $12.50; sixth meeting)	—
13	Saturday	£5 (US $6.25)	—	—
14	Sunday	£5 (US $6.25)	—	—
15	Monday	£5 (US $6.25)	£15 (US $18.75; seventh meeting)	£35 (US $43.75; seventh meeting)

[a]Total of £75 (US $93.76).

[b]Total of £70 (US $87.51).

[c]Total of £35 (US $43.75).

[d]Not applicable.

Measures

BACtrack Skyn

The TAS used in this study was the BACtrack Skyn (model T15). It was worn on the participants' preferred wrist, but they could change which wrist they wore it on during the study period if that was comfortable for them. The Skyn continuously measured the participant's TAC while being worn, as well as skin temperature (°C). Output was analyzed at 1-minute intervals. The participants could remove the TAS at any time if they did not wish to wear it and were required to remove it for bathing as it is not waterproof. The CM group were told that they could remove it once a day for up to 60 consecutive minutes to bathe and still be eligible for their CM reward. If it was removed for longer than an hour, they would no longer be able to receive the CM reward. If they wore the TAS according to this and the TAC did not increase above our set threshold of 115.660 µg/L, they were eligible for the CM reward for that day. Our set threshold was based on previous work by the research team [35].

TLFB Method

A TLFB was completed at meetings 2 to 7 to assess self-reported alcohol consumption and compare it against TAC. The TLFB is a calendar-based measure to record self-reported substance use. A day was considered from midnight to 11:59 PM.

Postwear Surveys

Participants completed a postwear survey on their experience of wearing the Skyn at their last meeting. This survey was adapted from the work by Alessi et al [20]. If participants were randomized to the CM group, they also completed a survey on their CM experience. This survey was adapted from the work by Miguel et al [62].

Outcomes

Feasibility of the Trial

This primary outcome was the feasibility of this study design. Feasibility was defined by enrollment, participation, device tampering, removals, adjustments, malfunction rates, and the number of TASs returned.

Enrollment

Participants who were identified, approached, eligible, and enrolled were recorded. Participants who were approached but did not wish to participate were asked about their main reason for this. Participant safety was recorded via adverse events for each participant. The number of participants enrolled on each service was also recorded.

Participation

The number of attended meetings by enrolled participants, withdrawal rate, compliance with wearing the TAS, and reasons for incomplete participant data (nonattendance, TAS data overwriting, and TAS technical faults) were recorded. Participants were asked what their main reason was for nonattendance or for not complying with TAS wear.

Removals

TAS removals were defined as a skin temperature of <30 °C for >2 minutes. Removals were recorded for each participant.

Tampering

Participants were asked about tampering with the TAS if there was no clear reason for missing TAC data in the output.

Malfunction

TAS errors, missing data due to a technical fault, and charging or syncing issues were recorded, as well as how much data were missing for these reasons.

TAS Return

The number of Skyn devices that were returned to the research team at the end of each participant's study period was recorded.

CM Delivery

The feasibility of TASs to measure CM target behavior was a combination of the factors mentioned previously, including compliance, removals, TAS tampering, malfunction, and accuracy.

Secondary Outcomes

Acceptability

This was measured using the postwear survey.

Accuracy

Accuracy outcomes were measured using the TLFB and TAC data. The TLFB was self-reported [63] and was used to determine alcohol-drinking days and how many units were consumed (in the United Kingdom, 8 g or 10 mL of pure ethanol=1 unit). The TAS (BACtrack Skyn) continuously measured TAC and skin temperature (°C), and from this, Skyn removal and then reported drinking days based on the TAC (µg/L) were determined. The TLFB and TAC data were analyzed to determine TAS accuracy compared to self-report data.

To note, the alcohol-drinking day defined using TAC is different to the CM intervention criteria, which allowed for a low amount of drinking and had a higher TAC criterion of 115.660 µg/L.

Data Handling

There is currently no guidance from BACtrack to determine drinking event criteria. Courtney et al [34] described their procedure for processing Skyn output to identify drinking episodes, and we used these guidelines when processing participant data. We replaced any negative values recorded with 0 in the data output. Missing data were classified as any minutes not reported. Removed data were those reporting a temperature of <30 °C for >2 minutes.

An alcohol event was based on TAC greater than a specific value (µg/L) for more than a set number of minutes. These criteria were chosen due to previous work carried out by this research team [35,59].

Analysis

Descriptive statistics were reported for baseline and demographic variables. All statistical analyses were conducted using SPSS (version 28; IBM Corp).

The feasibility outcomes reported included enrollment and recruitment rate, participant attendance, response and compliance, removals, TAS tampering and error, and TAS return. Appropriate summary statistics were reported, and independent-sample 2-tailed *t* tests were used to compare means between the control and CM groups when appropriate.

Summary statistics were reported for the secondary outcomes to be explored for a possible future larger trial. The postwear survey answers were reported. TAS accuracy was determined by analyzing the TAS data compared to self-report TLFB. The analysis focused on the sensitivity, specificity, positive predictive value, negative predictive value, and percentage accuracy in classification of TAC compared to TLFB as the gold standard. Recorded drinking and abstinent days were assessed using Spearman rank correlations. Sensitivity in detecting alcohol events and specificity in classifying an alcohol-drinking day versus a non–alcohol-drinking day were assessed using different TAC criteria: TAS 15 (TAC>15 µg/L for >15 min), TAS 60 (TAC>15 µg/L for >60 min), and TAS 90 (TAC>15 µg/L for >90 min).

Ethical Considerations

This study was approved by the Cornwall and Plymouth Research Ethics Committee (reference: 23/SW/0066) and registered on the Open Science Framework and International Standard Randomised Controlled Trial Number (reference: ISRCTN46845361). Informed written consent was obtained from all participants and their data were anonymized. All participants received a £5 (US $6.25) voucher at every meeting they attended (total for all 7 meetings=£35 [US $43.75]), a voucher for the return of the TAS on the last meeting (£10 [US $12.50]), and reimbursement for travel costs to each meeting.

Therefore, each participant could have received up to £45 (US $56.25) for their participation plus travel costs.

Results

Participants

A total of 32 healthy adult participants (n=10, 31% female and n=22, 69% male) enrolled, and a total of 29 (91%) completed the study. In total, 3 participants withdrew (n=2, 67% from the control group and n=1, 33% from the CM group). Of these 3 withdrawals, 2 (67%) happened during the study period (P8 [control group] on day 3 and P28 [CM group] on day 12), and 1 (33%) occurred at the end of the first meeting (control group) due to the specific type of voucher, which could not be used at the participant's local supermarket. P8 and P28 withdrew because of personal circumstances changing. Therefore, a total of 29 participants were included in the analysis (n=10, 34% female and n=19, 66% male; Figure 1). Table 3 shows the characteristics of the 29 participants included in the analysis split by group (control vs CM).

Figure 1. Flowchart of participant recruitment and retention. CM: contingency management.

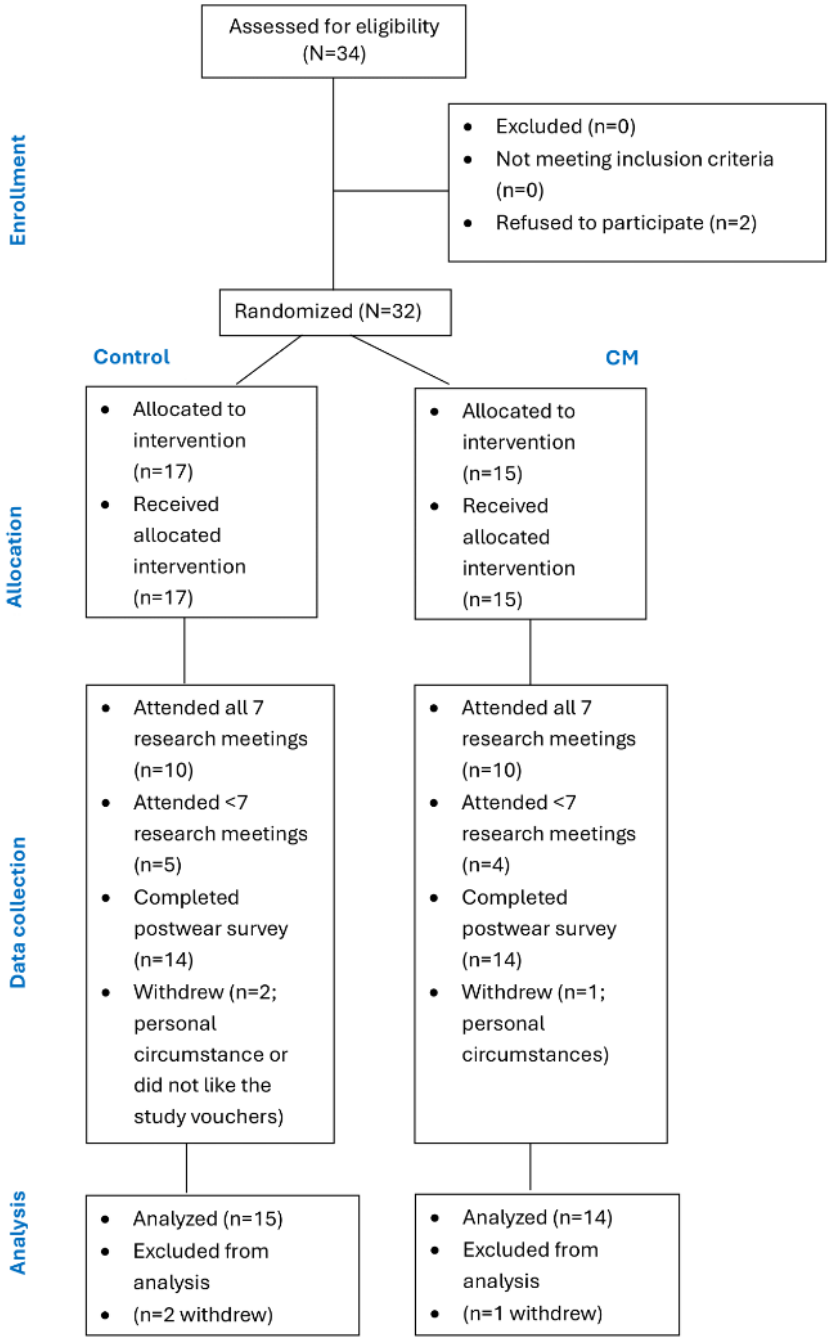


Table 3. Differences in group characteristics.

Characteristic	Control group (n=15)	CM[a] group (n=14)	Group intervention difference
Sex[b], n (%)			χ^2_1=0.4; *P*=.52
Female	6 (40)	4 (28.6)	
Male	9 (60)	10 (71.4)	
Age (y), mean (SD; range)	44.4 (13.78; 30-75)	51.5 (9.77; 35-60)	*P*=.12
BMI (kg/m^2), mean (SD; range)[c]	27.3 (6.24; 20.8-42.3)	28.7 (6.35; 22.2-42.5)	*P*=.58
Height (cm), mean (SD; range)	168.17 (6.40; 162-177)	171.29 (7.49; 162-185)	*P*=.24
Weight (kg), mean (SD; range)	77.2 (17.16; 53-112.5)	84.63 (19.75; 59-126.33)	*P*=.31
Ethnicity[b], n (%)			χ^2_{16}=43.3; *P*<.001
Black African	3 (20)	0 (0)	
Black British	2 (13.3)	0 (0)	
Caribbean	0 (0)	1 (7.1)	
Hispanic	1 (6.7)	0 (0)	
White British	7 (46.7)	9 (64.3)	
White Irish	1 (6.7)	3 (21.4)	
White Polish	0 (0)	1 (7.1)	
White Russian	1 (6.7)	0 (0)	
Units of alcohol consumed, mean (SD; range)	117.14 (105.83; 0-316.69)	53.33 (93.21; 0-294.69)	*P*=.14
Days in which alcohol was consumed (TLFB[d]), mean (SD)	8.14 (6.44)[e]	4.50 (6.62)[f]	*P*=.08
Average units consumed over days of alcohol consumption, mean (SD)	117.14[g] (105.83)	53.33[h] (93.21)	—[i]
TAS[j] removals (min), n (%)	60,879 (20.8)[k]	30,103 (11)[l]	*P*=.08
TAS missing data (min), n (%)	28,213 (9.6)[k]	28,591 (10.4)[l]	*P*=.80
Meetings attended, n (%)	94 (89.5)[m]	91 (92.9)[n]	*P*=.57
Participants who completed the postwear survey, n (%)	14 (93.3)	14 (100)	*P*=.34
Participants who returned the TAS, n (%)	14 (93.3)	14 (100)	*P*=.34

[a]CM: contingency management.

[b]Ethnicity is represented as the frequency of individuals in each group identifying as Black African, Black British, Caribbean, Hispanic, White British, White Irish, White Polish, and White Russian and was compared across sexes using a chi-square analysis.

[c]One female and one male participant requested not to be weighed, so BMI could not be calculated (both in the control group).

[d]TLFB: timeline followback.

[e]Total of 114 days.

[f]Total of 63 days.

[g]1640.01 units over 114 days.

[h]746.67 units over 63 days.

[i]No group comparison was possible.

[j]TAS: transdermal alcohol sensor.

[k]n=293,186.

[l]n=274,218.

[m]n=105.

[n]n=98.

Feasibility of the Trial

Enrollment

A total of 34 potentially eligible participants were approached. Only 6% (2/34) who were identified by staff declined to participate after discussing the study with the researcher. Their reasons were (1) that they did not like wearing a watch so did not think that they would like to wear a TAS and (2) that they were very busy and could not commit to the regular meetings. No adverse events occurred. The services involved were also willing to help with recruitment, and all recruited at least 4 participants over the 5-month recruitment period.

Participation

It was feasible to enroll more than our target sample size (30 participants) within 5 months. The first participant was enrolled on July 5, 2023, and the final participant (N=32) was enrolled on December 6, 2023. These data could be used to inform the sample size power calculation for a future larger trial.

Among the 29 participants who completed the study period and did not withdraw, a total of 203 meetings were arranged (29 × 7 visits). Of these 203 meetings, 185 (91.1%) were attended. Reasons for nonattendance included illness, their partner being ill, being double booked with another health appointment, a broken phone so the researcher could not be contacted, and hospitalization. The research team decided to book research meetings one at a time, so at meeting 1, meeting 2 was agreed upon, and so on. The day before each meeting, a reminder SMS text message was sent out with the time and date and asking the participants to let the research team know whether this needed to be changed. This design seems feasible to maintain a high attendance rate.

Only 1 of the 29 participants (3%) spoke to the research team about removing their TAS early and removed it at the sixth meeting (2 days early) because they did not want to wear it anymore as they felt that “it is messing with my head” (P22).

Among the 3 services involved, the Assertive Alcohol Outreach Team recruited 53% (17/32) of the participants, the Pier Road Project recruited 34% (11/32) of the participants, and Wandsworth Community Drug and Alcohol Service recruited 12% (4/32) of the participants.

Removals

We defined TAS removal as >2 minutes during which the skin temperature was of <30 °C. We reminded participants to wear the TAS as much as possible, including while asleep. However, the TAS is removable by the wearer, and they could choose not to wear it if they wanted to and must remove it for water activities (bathing and swimming). Of the total minutes collected from the 29 participants (464,324 minutes of data, approximately 322 days), there was a total of 19.59% (90,982/464,324) of minutes of removals (approximately 63 days).

The control group had a total of 20.21% (60,879/301,179) of minutes of removals. The CM group had a total of 10.79% (30,103/278,862) of minutes of removals (t_{27}=1.843; P=.08; d=0.685).

Tampering

One participant (P1) admitted to turning the Skyn off while wearing it when they did not want to be monitored. No other evidence or reports of tampering from any other participant was recorded. This participant had 65.65% (13,173/20,067) of minutes of their participation time successfully recorded and 34.35% (6893/20,067) of missing data. Of these missing data, approximately half could be due to not attending meetings and data being overwritten (3386/6893, 49.12%). However, all other meetings were attended, so the 50.88% (3507/6893) of minutes of missing data could be due to the TAS being turned off or technology error and additional data overwriting (approximately 2.5 days’ worth of data).

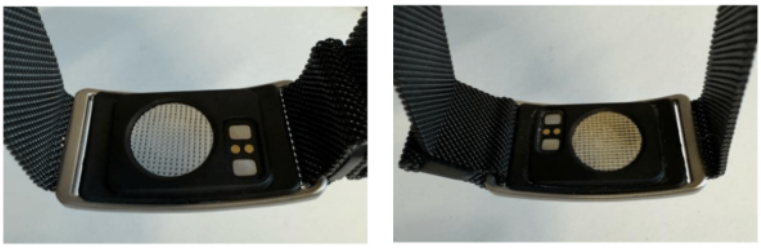
Malfunction

A total of 16 Skyns were ordered in phases during the study period. The first order of 6 Skyns had 2 issues: one Skyn had trouble pairing for the first 2 weeks, and another would not charge or pair with any research iPhone. BACtrack sent 3 Skyn replacements, and 2 of these would not pair with a research iPhone at first (they did 1 month later). We ordered a second batch of 7 Skyns halfway through data collection, and there were no issues with any devices from this order. BACtrack were unsure about the reasons for the issue with pairing and charging. Other issues experienced included syncing taking >40 minutes on occasions (the usual time is approximately 5 minutes). This was most likely due to poor connection or Wi-Fi.

Apart from 1 Skyn that arrived and could not be charged, there were no charging issues. The Skyn website states that the battery lasts for approximately 10 days; however, we found that it lasted for >2 weeks. Participants were provided with a fully charged Skyn at the first meeting. There were 2 participants whose battery decreased to a low level within their 2-week study period, so the warning red light started flashing. As the TASs were typically able to retain the charge for the entire study period, we did not provide the charging cable to participants. If the TASs will be used for >2 weeks in a future study, either the charging cable will need to be provided or there will need to be planned charging time with the researchers.

There was evidence of Skyn degradation after 2 months, which is partly why it was decided to order another batch of Skyns halfway through data collection [33]. Another reason for this order was their state of cleanliness (Figure 2). BACtrack do provide guidelines on how to clean the sensor; however, after following these instructions, the sensor could not be cleaned well.

Figure 2. These 2 photos of the BACtrack Skyn sensor show a brand-new, unworn Skyn (left) and a worn Skyn (right).



TAS Return

All but 1 Skyn were returned. This was due to the participant disengaging from the service and not being contactable by the researcher or key worker. One other participant did not attend their last research meeting, but they arranged an additional meeting to return the TAS.

CM and Intervention Delivery

CM rewards were delivered if 2 criteria were met: (1) the TAC for the day (midnight-11:59 PM) had a peak of <115.660 µg/L and (2) there were no removals lasting >60 minutes (temperature of <30 °C).

The TAS data were downloaded at each meeting (meetings 2 to 7) and checked, and any CM vouchers were provided for the previous 2 to 3 days. This was considered the best way to implement CM as close to those days in which the behavior occurred as possible. Table 4 details the amount of CM rewards earned by each participant and any reason why CM rewards were not earned.

Table 4. Contingency management (CM) rewards earned by participants in the CM group.

Participant ID	Days in which CM rewards were earned[a]	Total amount earned[b]	Days in which CM rewards were not earned[c]	Reason why CM rewards were not earned[d]
P4	0	£0	15	15 days=TAC[e] over the limit
P5	13	£105 (US $131.26)	2	1 day=TAC over the limit; 1 day=removal for >60 min
P6	12	£95 (US $118.76)	3	1 day=TAC over the limit; 2 days=removal for >60 min
P7	10	£85 (US $106.26)	5	1 day=TAC over the limit; 4 days=removal for >60 min
P10	12	£90 (US $112.51)	3	1 day=TAC over the limit; 2 days=removal for >60 min
P11	0	£0	15	15 days=TAC over the limit
P16	2	£10 (US $12.50)	13	6 days=removal for >60 min; 7 days=DNA[f] appointment and data overwritten
P17	5	£25 (US $31.25)	10	6 days=TAC over the limit; 2 days=removal for >60 min; 2 days=DNA appointment and data overwritten
P18	9	£55 (US $68.75)	6	6 days=removal for >60 min
P19	9	£75 (US $93.76)	6	3 days=removal for >60 min; 3 days=DNA appointment and data overwritten
P21	13	£110 (US $137.51)	2	2 days=removal for >60 min
P22	1	£5 (US $6.25)	11	9 days=TAC over the limit; 2 days=removal for >60 min; handed TAS[g] back 2 days early
P25	3	£15 (US $18.75)	12	2 days=TAC over the limit; 10 days=removal for >60 min
P29	11	£85 (US $106.26)	4	4 days=removal for >60 min

[a]Total: 48.3% (100/207) of days.

[b]Total: £755 (US $943.82).

[c]Total: 51.7% (107/207) of days.

[d]Total: 24.6% (51/207) of days with transdermal alcohol concentration over the limit, 21.3% (44/207) of days of removals, and 5.8% (12/207) of days of data overwritten.

[e]TAC: transdermal alcohol concentration.

[f]DNA: did not attend.

[g]TAS: transdermal alcohol sensor.

Control group participants attended 89.5% (94/105) of the meetings, and the CM group participants attended 92.9% (91/98) of the meetings (t_{27}=−0.573; *P*=.57; *d*=−0.213). The control group self-reported 48.6% (107/220) of abstinent days, and the CM group self-reported 70% (147/210) of abstinent days (t_{27}=−1.403; *P*=.17; *d*=−0.522). The proportion of days in which the Skyn data reported <1 hour of removals was 15.9% (35/220) for the control group and 24.8% (52/210) for the CM group (t_{27}=−1.326; *P*=.20; *d*=−0.493). The difference in the amount of units consumed between groups was not statistically significant (control: mean 109.33, SD 27.47; CM: mean 53.33, SD 24.91; t_{27}=1.503; *P*=.14; *d*=0.559).

Acceptability and Postwear Survey

A total of 97% (28/29) of the participants completed the postwear survey. One participant did not complete it because they did not attend their final research meeting and became unreachable. When asked about physical comfort, most participants (11/28, 39%) rated it to be quite to very comfortable, with the average score being 8.57/10 (SD 0.71; 10=very comfortable). Social comfort (how they felt wearing it in public) was also rated highly, with an average score of 9.63/10 (SD 0.00; 10=very socially comfortable). When asked how often they noticed the Skyn on their wrist, only 11% (3/28) of the participants said that they never noticed it when wearing it. Most (18/28, 64%) noticed it once or twice a day to every hour, but 25% (7/28) did report noticing it several times per hour. When asked to rate its interference with various activities

(exercise, mood, normal work, enjoyment of life, ability to concentrate, social life, and clothing choices), these were all rated with an average score of <2 (1=no interference at all; general activity: 1.75, SD 0.00; exercise: 1.07, SD 0.00; mood: 1.11 SD 0.00; work: 1.07, SD 0.00; enjoyment of life: 1.21, SD 0.00, ability to concentrate: 1.43, SD 0.00; social life: 1, SD 0.00; choice of clothing: 1.24, SD 0.00). The only activity that ranked higher was sleeping, with an average score of 2.71 (SD 2.83; 1=no interference at all).

A total of 32% (9/28) of the participants reported experiencing a mark on their skin from the TAS, with the other 68% (19/28) reporting never experiencing a mark or side effect. These side effects included itching (mean 1.95, SD 0.00), sweating (mean 1.88, SD 0.00), soreness (mean 1.07, SD 0.00), and irritation (mean 1.50, SD 0.00), with the scale being from 1=not noticeable to 10=unbearable. Participants were asked whether they would continue to wear the TAS for longer than the 2-week study period, and 96% (27/28) reported *yes*, and only 4% (1/28) reported *no*.

The following statements—"the device is too uncomfortable to wear for any longer," "I wouldn't want to wear it any longer because I am embarrassed," and "I want to wear short sleeves but won't while wearing the device"—were answered by all participants with "not true." Statements asking about changing their clothing choices, wanting to remove the TAS, and not liking the regular download visits all had an average score of 0.25 (SD 0.00 for all these statements) (on a scale of 0-4 where 0=*not true* and 4=*very true*). Only 4% (1/28) of the participants stated that they were tired of explaining the device to people. In total, 4% (1/28) stated that they were "ready to stop wearing the device because I am just ready to be done," and another participant (1/28, 4%) said that the "financial compensation for wearing the device any longer than this would not be worth it." A total of 7% (2/28) of the participants said that it was true that "I would not continue wearing the device because I do not like having to do the downloads at a specific time."

When asked about their drinking, 11% (3/28) self-reported that they were able to completely reduce their drinking, 25% (7/28) reported that they reduced it quite a bit, 14% (4/28) reported that they reduced it somewhat, 7% (2/28) reported that they tried but did not reduce it, and 11% (3/28) reported that they did not reduce or try to reduce it. In total, 32% (9/28) of the participants were abstinent throughout their study period.

Many participants (12/28, 43%) liked the in-person vouchers that were given at each meeting; however, 7% (2/28) would have preferred an e-voucher, 21% (6/28) would have preferred a bank transfer, and 18% (5/28) would be happy with any of these options (in-person voucher, e-voucher, or bank transfer). In total, 7% (2/28) reported preferring only an e-voucher or bank transfer, and 4% (1/28) preferred an in-person voucher or a bank transfer. Finally, participants were asked about the ease of meeting the researcher for the meetings, and 89% (25/28) reported that it was *not at all difficult*, whereas the other 11% (3/28) reported that it was *a little bit difficult*.

CM Survey

The 48% (14/29) of participants who were allocated the CM group completed another survey specifically on their experience of the CM rewards. All participants in the CM group (14/14, 100%) completed this survey. When asked about how easy it was to understand the CM reward procedure, most (9/14, 64%) said that it was very easy, but 21% (3/14) rated it as very difficult. When asked whether the CM had any effect on their response to treatment, half (7/14, 50%) said that it helped them either a lot or a little, 43% (6/14) said that it did not make a difference, and 7% (1/14) said that it had a negative impact. When asked whether they liked the CM rewards, all (14/14, 100%) said that they liked them either a lot or a little. When asked when CM rewards would help other people who seek substance treatment for alcohol dependence, 86% (12/14) said that it would help them a lot or a little, and 14% (2/14) said that it would depend on the person.

Accuracy

Participants wore the TAS for a cumulative total of 580,040 minutes. A total of 80.05% (464,324/580,040) of minutes of this participation was successfully recorded by the TAS. Of these 464,324 minutes of recorded data, there was a total of 90,982 (19.59%) minutes that suggested removal, a total of 56,858 (12.25%) minutes that were missing for unknown reasons, and a total of 58,986 (12.7%) minutes that were missing due to being overwritten as a result of not attending appointments. The Skyn is not waterproof, so participants were required to remove it for showering, baths, and any other water activities (eg, swimming). No fit adjustments were required.

The TAS recorded data for a total of 388 days. Of these 388 days, 337 (86.9%) had >60 minutes of removals or missing data. There was a total of 68.6% (266/388) of days in which there were >300 minutes of removed or missing data (not necessarily a consecutive 5-hour period).

TLFB Versus Skyn TAC

Overview

TLFB drinking days were recorded via the participant reporting any alcohol consumed. TAS drinking days were recorded according to the TAC criteria. We considered 3 different criteria for detecting alcohol-drinking days: TAC of >15 µg/L for >15 minutes, TAC of >15 µg/L for >60 minutes, and TAC of >15 µg/L for >90 minutes. The agreement between the TLFB and these criteria is reported in Table 5.

Table 5. Timeline followback (TLFB)– and transdermal alcohol sensor (TAS)–reported alcohol-drinking and nondrinking days.

Criteria	Alcohol-drinking days detected	Alcohol-drinking days in agreement (n=388), n (%)	Non–alcohol-drinking days in agreement (n=388), n (%)	Days reported by the TAS but not by TLFB as alcohol-drinking days (n=388), n (%)	Days reported by TLFB but not by the TAS as alcohol-drinking days (n=388), n (%)
TAC[a] >15 µg/L for >15 min		164 (42.3)	145 (37.4)	65 (16.8)	14 (3.6)
TLFB	185				
TAS	227				
TAC>15 µg/L for >60 min		146 (37.6)	171 (44.1)	39 (10.1)	32 (8.2)
TAS	185				
TLFB	185				
TAC>15 µg/L for >90 min		139 (35.8)	179 (46.1)	31 (8)	39 (10.1)
TAS	185				
TLFB	170				

[a]TAC: transdermal alcohol concentration.

Criteria for a Drinking Event of TAC>15 µg/L for >15 Minutes

The TLFB and TAS agreed on 42.3% (164/388) of days as alcohol-drinking days and on 37.4% (145/388) of days as abstinent days. There were 3.6% (14/388) of days in which the TLFB reported an alcohol-drinking day and the TAS did not and 16.8% (65/388) of days in which the TAS reported an alcohol-drinking day and the TLFB did not.

When splitting the groups and conducting a correlation between TAS- and TLFB-reported drinking days, both were positively significantly correlated, but the CM group had a stronger correlation effect (control group: r_{15}=0.625 and P=.01; CM group: r_{14}=0.836 and P<.001).

Criteria for a Drinking Event of TAC>15 µg/L for >60 Minutes

The TLFB and TAS agreed on 37.6% (146/388) of days as alcohol-drinking days and on 44.1% (171/388) of days as abstinent days. There were 8.2% (32/388) of days in which the TLFB reported an alcohol-drinking day and the TAS did not and 10.1% (39/388) of days in which the TAS reported an alcohol-drinking day and the TLFB did not.

When splitting the groups and conducting a correlation between TAS- and TLFB-reported drinking days, both were positively significantly correlated, but the CM group had a stronger correlation effect (control group: r_{15}=0.764 and P<.001; CM group: r_{14}=0.895 and P<.001).

Criteria for a Drinking Event of TAC>15 µg/L for >90 Minutes

The TLFB and TAS agreed on 35.8% (139/388) of days as alcohol-drinking days and on 46.1% (179/388) of days as abstinent days. There were 10.1% (39/388) of days in which the TLFB reported an alcohol-drinking day and the TAS did not and 8% (31/388) of days in which the TAS reported an alcohol-drinking day and the TLFB did not.

When splitting the groups and conducting a correlation between TAS- and TLFB-reported drinking days, both were positively significantly correlated, but the CM group had a stronger correlation effect (control group: r_{15}=0.805 and P<.001; CM group: r_{14}=0.913 and P<.001).

Skyn TAC

Reasons why the TAS did not detect an alcohol event but it was reported in the TLFB could include a large amount of missing data; TAS removal; a low amount of alcohol self-reported; or it being the last day of participation, which may mean that there was not enough time for alcohol to appear in the TAC output. There were other days in which there was no obvious reason why the TAS did not detect alcohol consumption.

Reasons why the TAS reported an alcohol-drinking day but the participant did not self-report alcohol consumption could include a sudden TAC spike, which suggests contact with an alcohol-containing product (eg, aftershave, deodorant, or hand sanitizer); the skin temperature suggesting that the TAS was not being worn at the time of TAC event detection, so the TAS may have been on a table near an alcohol spill or alcohol-containing product; and the TAS appearing to detect an alcohol event, so the self-report was inaccurate. There were other days in which there was no obvious reason why the participant did not self-report alcohol consumption.

In Figure 3, a visual presentation of the data for each TAC criteria is provided. It shows TLFB and the TAS criteria of TAS 15 (TAC>15 µg/L for >15 min), TAS 60 (TAC>15 µg/L for >60 min), and TAS 90 (TAC>15 µg/L for >90 min). The presence of the color corresponding to each of these TAC criteria indicates that an alcohol-drinking day was detected.

Figure 3. Visual representation of participants' Timeline Followback (TLFB)– and transdermal alcohol sensor (TAS)–detected alcohol-drinking days for (A) the control group and (B) the contingency management (CM) group. The figure includes 3 different TAS criteria for detecting a drinking day: TAS 15 (transdermal alcohol concentration [TAC]>15 µg/L for >15 min), TAS 60 (TAC>15 µg/L for >60 min), and TAS 90 (TAC>15 µg/L for >90 min).

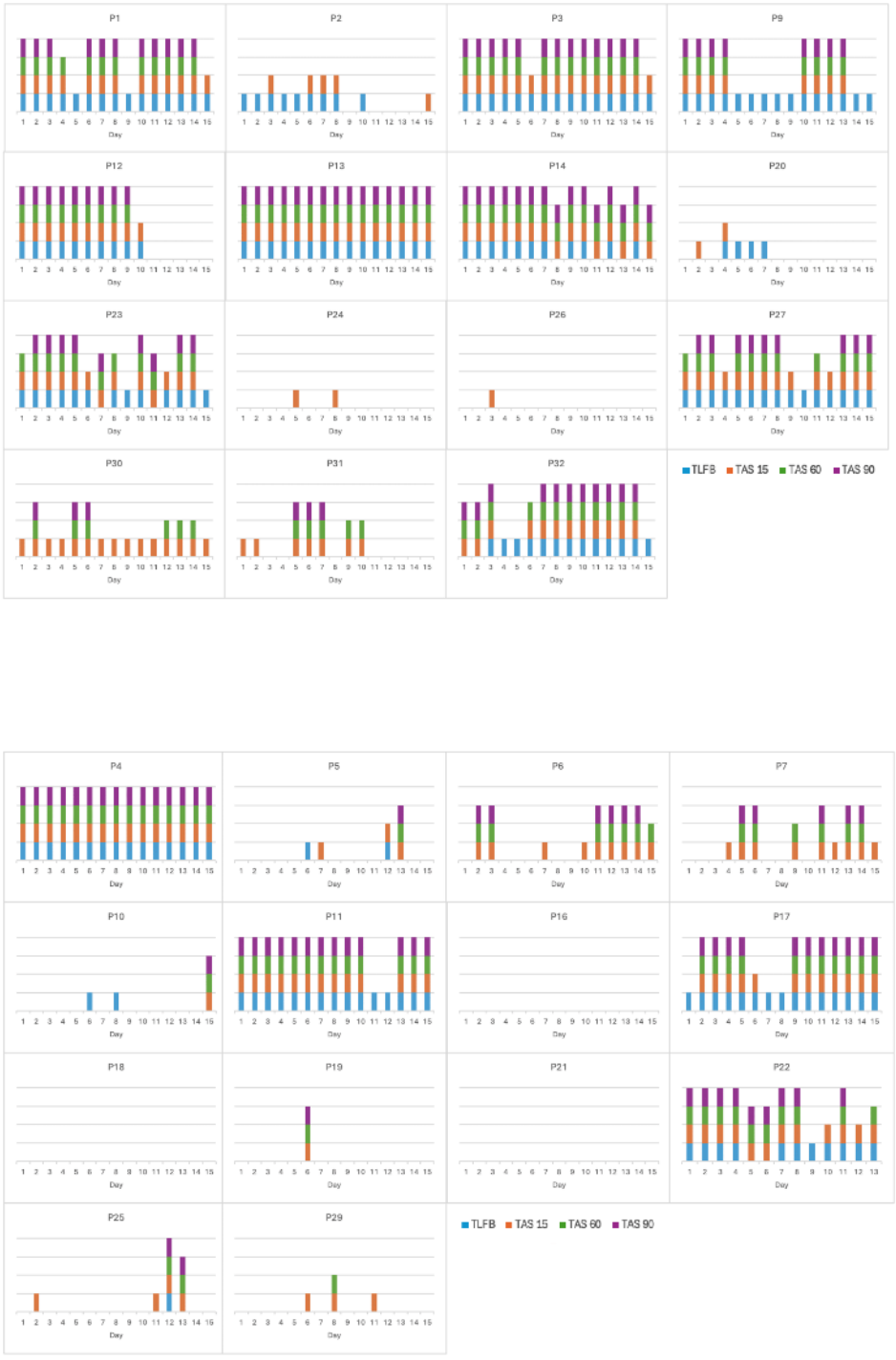
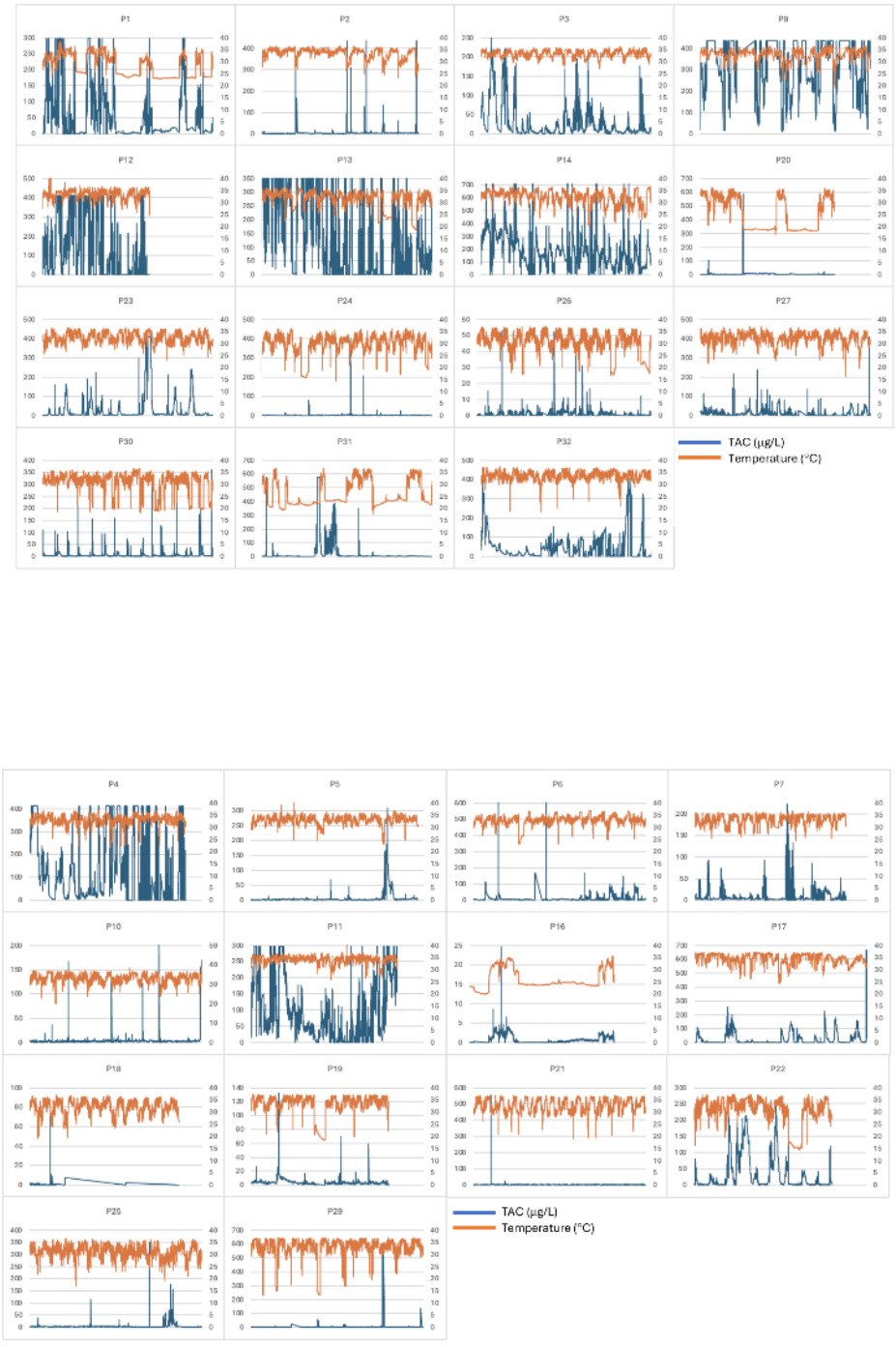


Figure 4 shows the data for TAC and the skin temperature of all participants who wore the TAS for the entire study period. There were participants whose TAC (blue line) peaked on separate occasions, suggesting multiple alcohol events, for example, P5 or P17. In contrast, other participant data suggest more frequent alcohol consumption, for example, P9 and P13. The orange lines depict the skin temperature recorded; the skin temperature for P1 and P20 suggests regular, long TAS removal.

Figure 4. Individual participant Skyn data (excluding withdrawn participants). The primary axis shows the transdermal alcohol concentration (TAC; µg/L) data (blue line), and the secondary axis shows the skin temperature (°C) data (orange line).



Sensitivity, Specificity, and Correlations

We calculated the sensitivity, specificity, positive predictive value, negative predictive value, and percentage accuracy in classification for 3 drinking event thresholds: TAS 15 (TAC>15 µg/L for >15 min), TAS 60 (TAC>15 µg/L for >60 min), and TAS 90 (TAC>15 µg/L for >90 min), shown in Table 6.

Table 6. Transdermal alcohol concentration (TAC) criteria—sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and percentage accuracy classification (PAC).

Criteria	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	PAC
TAC>15 µg/L for 15 min	92.7	68.29	71.74	91.5	0.80
TAC>15 µg/L for 60 min	82.49	80.5	78.92	83.85	0.81
TAC>15 µg/L for 90 min	78.09	85.31	81.76	82.19	0.82

A Spearman correlation found a significant relationship between TLFB- and Skyn-measured alcohol-drinking days at the TAC>15 µg/L for >15 minutes threshold (r_s=0.695, bias–corrected and accelerated [BCa] 95% CI 0.417-0.883; *P*<.001) [29]. A Spearman correlation between the number of standard alcohol units self-reported and the average TAC values was also found to be significant (r_s=0.770, BCa 95% CI 0.632-0.852; *P*<.001) [29].

A Spearman correlation found a significant relationship between TLFB- and Skyn-measured alcohol-drinking days at the TAC>15 µg/L for >60 minutes threshold (r_s=0.773, BCa 95% CI 0.599-0.861; *P*<.001) [29]. A Spearman correlation between the total number of standard alcohol units self-reported and the TAS-reported drinking days at the TAC>15 µg/L for >60 minutes threshold was also found to be significant (r_s=0.740, BCa 95% CI 0.439-0.891; *P*<.001) [29].

A Spearman correlation found a significant relationship between TLFB- and Skyn-measured alcohol-drinking days at the TAC>15 µg/L for >90 minutes threshold (r_s=0.784, BCa 95% CI 0.625-0.868; *P*<.001) [29]. A Spearman correlation between the total number of standard alcohol units self-reported and the TAS-reported drinking days at the TAC>15 µg/L for >90 minutes threshold was also found to be significant (r_s=0.745, BCa 95% CI 0.460-0.892; *P*<.001) [29].

Discussion

Principal Findings

This study aimed to explore the feasibility, strengths, and limitations of using a TAS to monitor alcohol consumption in individuals in treatment for AUD with or without CM to promote abstinence or low-level alcohol consumption. The findings suggest that TAS-delivered CM to encourage alcohol reduction was feasible and acceptable to participants. There was a high correlation between TAS-recorded alcohol-drinking days and self-reported alcohol units, suggesting that the TAS was accurate in measuring the desired behavior. There was also high recruitment, attendance, and compliance of participants. Good rates of meeting attendance and data completeness were achieved. It was feasible to deliver CM using the TAS; however, key features of this process were identified for improvement. This is the first study of TAS-delivered CM in the United Kingdom and the first targeted at alcohol service users for alcohol treatment and reduction.

Participants found the wear of the TAS acceptable and comfortable even for social occasions. They found that it did not interfere with daily activities, and most (19/28, 68%) did not experience any side effects (such as a rash or skin irritation) from the strap. There was a high attendance rate and willingness to participate from those approached. Participants liked the reminder SMS text messages about each meeting, and there were few issues when arranging meetings. Only 3% (1/29) of the participants lost contact with the researcher during participation and did not return the TAS.

However, the answers to the CM survey suggest that this process may need to be improved. Some participants (3/28, 11%) reported it as difficult to understand, and half (7/14, 50%) said that it made no difference or did not help their treatment. These findings suggest that there should be more consideration on how the CM procedure is described and presented to participants at the start. Most (12/14, 86%) said that they did think that CM rewards would help other people who seek treatment for alcohol dependence a little to a lot, with the other 14% (2/14) of the participants saying that it would probably help but only if the patient was motivated.

To deliver the CM, TAS data were downloaded and reviewed at meetings 2 to 7. This was feasible; however, there were issues that arose that would need to be dealt with before scaling up the study size. Considering other options for automated TAC data interpretation will reduce the time-consuming manual data checking, which would only increase the researchers' time burden if conducted on a larger trial. In addition, creating an automated system could increase consistency and avoid human error.

The recruitment and follow-up rates from this study were the same as or higher than the recruitment and follow-up rates of other TAS feasibility or pilot studies. Previous studies using TASs had sample sizes for analysis of 5 to 13 participants [2,12,34,64]. In total, 2 of these studies used the SCRAM monitor [2,64], and 2 used the BACtrack Skyn [12,34], but 3 [2,34,64] mentioned discomfort with the TAS as key feedback, with Courtney et al [34], who used the BACtrack, having one participant drop out due to discomfort.

Overall, the findings support the feasibility of implementing and delivering CM using a TAS over a 2-week wear period with alcohol service users currently receiving alcohol treatment. The TAS used, BACtrack Skyn, was well liked and rated as comfortable by participants, with little daily interference. The recruitment rates were high and are encouraging for a larger trial. TAS delivery of CM has been shown to be effective in the United States using the SCRAM monitor [1,2,26-31,65], but this pilot shows promise for translation to other TAS brands delivering CM in the United Kingdom, with these findings highlighting aspects to address and improve before scaling to a larger study design. On the basis of these results, future work could explore the possibility of solutions to these challenges and a cost-effectiveness assessment.

In the future, we would also recommend collecting initial data from the participants to determine each participant's baseline skin temperature. This could be done with the researcher present to ensure correct TAS wear. Some participants were observed by the researcher to be wearing the Skyn during the meeting but the Skyn temperature was reported as <30 °C. Therefore, it may be better to establish a baseline for each participant to then use for their Skyn data for better accuracy at determining wear and removals.

Strengths and Limitations

The strengths of this study are that it was able to demonstrate the feasible use of the BACtrack Skyn over 2 weeks with individuals currently diagnosed with alcohol dependence and receiving alcohol treatment. Participants wore the TAS in their natural settings and unsupervised, consistent with how it should be worn. While the objective of this study was to assess the feasibility of a larger trial, the data collected as part of this study were able to provide more evidence of how this population wears, uses, and experiences a TAS. There was a high meeting attendance and TAS return rate for this study and no issues with participants being unsure or needing additional training for using the TAS after the baseline training. This study continues to support the use of TASs among the population.

The limitations of this study are similar to those of the design of all the studies conducted as part of this PhD—participants were only recruited if they were willing to wear the TAS from the start of the study period. This means that no participants were recruited who were not willing to attempt wearing the TAS. While this is not a limitation in some considerations as the TAS would be a voluntary treatment option to service users if implemented in services, it does mean that the postwear survey findings may be skewed more positively. Only those who were interested in and willing to wear the TAS had the chance to complete the postwear survey at the end of the 2 weeks. However, participants being willing to wear the TAS for the study does not necessarily mean that they were positive about the technology or that they had a good experience of wearing it; therefore, while we note that the study design does exclude those who are not initially willing to attempt wearing a TAS, this may reflect a truer view of service users who would try wearing the TAS as part of alcohol treatment if TASs were to be implemented in clinical settings.

Conclusions

To conclude, if planning a future larger trial for TAS delivery of CM, the proposed design changes from this study are (1) changes to how TAC data are checked for CM rewards from a manual to an automatic process, (2) a clearer explanation of the CM procedure, (3) consideration of the times available for participant meetings to improve participant availability and reduce missed appointments, (4) consideration of other TAS brands that have longer data storage, (5) consideration of whether participants should be provided with an iOS device and trained to sync their data or whether participants should have this option with their personal iOS device if using the BACtrack Skyn (which has been deemed feasible by wearers), and (6) consideration of the assessment of each participant's baseline skin temperature. These points could be assessed via an internal pilot study.

Acknowledgments

This is a summary of independent research funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration South London. The views expressed are those of the authors and not necessarily those of the National Health Service (NHS), the NIHR, or the Department of Health and Social Care. CD and PD were supported by the NIHR Maudsley Biomedical Research Centre at the South London and Maudsley NHS Foundation Trust and King's College London. They were also supported by the NIHR Collaboration for Leadership in Applied Health Research and Care at King's College Hospital NHS Foundation Trust and the NIHR Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust. CD was supported by an NIHR Senior Investigator Award. SP has been partly funded by income from research grants obtained from Mundipharma Research Ltd and Camurus AB. SP was a Research Fellow at King's College London at the time of writing. The authors would also like to acknowledge Andrea Solar-Reyes and Noel Kehoe for their data collection work alongside EB for this study and Simon Coulton for his help in participant randomization.

Authors' Contributions

EB led on the conceptualization, methodology, analysis, investigation, resources, data curation, project administration, visualization, and drafting and revision of the manuscript. CD supported the conceptualization, methodology, analysis, supervision, and revision of the manuscript. SP supported the supervision and revision of the manuscript. PD supported the conceptualization, methodology, analysis, supervision, and revision of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[DOC File , 227 KB - humanfactors_v12i1e64664_app1.doc]

References

1. Barnett NP, Celio MA, Tidey JW, Murphy JG, Colby SM, Swift RM. A preliminary randomized controlled trial of contingency management for alcohol use reduction using a transdermal alcohol sensor. Addiction 2017 Jun 22;112(6):1025-1035 [FREE Full text] [doi: 10.1111/add.13767] [Medline: 28107772]
2. Barnett NP, Tidey J, Murphy JG, Swift R, Colby SM. Contingency management for alcohol use reduction: a pilot study using a transdermal alcohol sensor. Drug Alcohol Depend 2011 Nov 01;118(2-3):391-399 [FREE Full text] [doi: 10.1016/j.drugalcdep.2011.04.023] [Medline: 21665385]
3. Rash CJ, Petry NM, Alessi SM, Barnett NP. Monitoring alcohol use in heavy drinking soup kitchen attendees. Alcohol 2019 Dec;81:139-147 [FREE Full text] [doi: 10.1016/j.alcohol.2018.10.001] [Medline: 30308286]
4. Karns-Wright TE, Dougherty DM, Hill-Kapturczak N, Mathias CW, Roache JD. The correspondence between transdermal alcohol monitoring and daily self-reported alcohol consumption. Addict Behav 2018 Oct;85:147-152 [FREE Full text] [doi: 10.1016/j.addbeh.2018.06.006] [Medline: 29910035]
5. Barnett NP, Meade EB, Glynn TR. Predictors of detection of alcohol use episodes using a transdermal alcohol sensor. Exp Clin Psychopharmacol 2014 Feb;22(1):86-96 [FREE Full text] [doi: 10.1037/a0034821] [Medline: 24490713]
6. Dougherty DM, Charles NE, Acheson A, John S, Furr RM, Hill-Kapturczak N. Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking. Exp Clin Psychopharmacol 2012 Oct;20(5):373-381 [FREE Full text] [doi: 10.1037/a0029021] [Medline: 22708608]
7. Fairbairn CE, Kang D. Temporal dynamics of transdermal alcohol concentration measured via new-generation wrist-worn biosensor. Alcohol Clin Exp Res 2019 Oct 30;43(10):2060-2069 [FREE Full text] [doi: 10.1111/acer.14172] [Medline: 31469451]
8. Hill-Kapturczak N, Lake SL, Roache JD, Cates SE, Liang Y, Dougherty DM. Do variable rates of alcohol drinking alter the ability to use transdermal alcohol monitors to estimate peak breath alcohol and total number of drinks? Alcohol Clin Exp Res 2014 Oct 21;38(10):2517-2522 [FREE Full text] [doi: 10.1111/acer.12528] [Medline: 25335857]
9. Bond JC, Greenfield TK, Patterson D, Kerr WC. Adjustments for drink size and ethanol content: new results from a self-report diary and transdermal sensor validation study. Alcohol Clin Exp Res 2014 Dec 07;38(12):3060-3067 [FREE Full text] [doi: 10.1111/acer.12589] [Medline: 25581661]
10. Karns-Wright TE, Roache JD, Hill-Kapturczak N, Liang Y, Mullen J, Dougherty DM. Time delays in transdermal alcohol concentrations relative to breath alcohol concentrations. Alcohol Alcohol 2017 Jan 13;52(1):35-41 [FREE Full text] [doi: 10.1093/alcalc/agw058] [Medline: 27522029]
11. Swift RM, Martin CS, Swette L, LaConti A, Kackley N. Studies on a wearable, electronic, transdermal alcohol sensor. Alcohol Clin Exp Res 1992 Aug 11;16(4):721-725. [doi: 10.1111/j.1530-0277.1992.tb00668.x] [Medline: 1530135]
12. Rosenberg M, Ludema C, Kianersi S, Luetke M, Jozkowski K, Guerra-Reyes L, et al. Wearable alcohol monitors for alcohol use data collection among college students: feasibility and acceptability. Alcohol 2023 Sep;111:75-83. [doi: 10.1016/j.alcohol.2023.05.007] [Medline: 37295566]
13. Wang Y, Fridberg DJ, Shortell DD, Leeman RF, Barnett NP, Cook RL, et al. Wrist-worn alcohol biosensors: applications and usability in behavioral research. Alcohol 2021 May;92:25-34 [FREE Full text] [doi: 10.1016/j.alcohol.2021.01.007] [Medline: 33609635]
14. Fairbairn CE, Kang D, Bosch N. Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory. Drug Alcohol Depend 2020 Nov 01;216:108205 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108205] [Medline: 32853998]
15. Alessi SM, Barnett NP, Petry NM. Objective continuous monitoring of alcohol consumption for three months among alcohol use disorder treatment outpatients. Alcohol 2019 Dec;81:131-138 [FREE Full text] [doi: 10.1016/j.alcohol.2019.01.008] [Medline: 30710610]
16. Roache JD, Karns TE, Hill-Kapturczak N, Mullen J, Liang Y, Lamb RJ, et al. Using transdermal alcohol monitoring to detect low-level drinking. Alcohol Clin Exp Res 2015 Jul 18;39(7):1120-1127 [FREE Full text] [doi: 10.1111/acer.12750] [Medline: 25988708]
17. Simons JS, Wills TA, Emery NN, Marks RM. Quantifying alcohol consumption: self-report, transdermal assessment, and prediction of dependence symptoms. Addict Behav 2015 Nov;50:205-212 [FREE Full text] [doi: 10.1016/j.addbeh.2015.06.042] [Medline: 26160523]
18. Brobbin E, Deluca P, Hemrage S, Drummond C. Acceptability and feasibility of wearable transdermal alcohol sensors: systematic review. JMIR Hum Factors 2022 Dec 23;9(4):e40210 [FREE Full text] [doi: 10.2196/40210] [Medline: 36563030]
19. Brobbin E, Deluca P, Hemrage S, Drummond C. Accuracy of wearable transdermal alcohol sensors: systematic review. J Med Internet Res 2022 Apr 14;24(4):e35178 [FREE Full text] [doi: 10.2196/35178] [Medline: 35436239]
20. Alessi SM, Barnett NP, Petry NM. Experiences with SCRAMx alcohol monitoring technology in 100 alcohol treatment outpatients. Drug Alcohol Depend 2017 Sep 01;178:417-424 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.05.031] [Medline: 28709081]
21. Caluzzi G, Pennay A, Cook M, Wright C, Norman T, Kuntsche E. Transdermal monitors to assess alcohol consumption in real-time and real-life – a qualitative study on user-experience. Addict Res Theory 2019 Jan 09;27(4):354-361. [doi: 10.1080/16066359.2018.1530765]

22. Croff JM, Hartwell ML, Chiaf AL, Crockett EK, Washburn IJ. Feasibility and reliability of continuously monitoring alcohol use among female adolescents and young adults. Drug Alcohol Rev 2021 Nov 17;40(7):1143-1154 [FREE Full text] [doi: 10.1111/dar.13045] [Medline: 32185847]
23. Goodall CA, Neville FG, Williams DJ, Donnelly PD. Preliminary research informing policy on remote alcohol monitoring in criminal justice: the Scottish experience. Int J Public Health 2016 Nov 13;61(8):865-872. [doi: 10.1007/s00038-016-0886-9] [Medline: 27624623]
24. Neville FG, Williams DJ, Goodall CA, Murer JS, Donnelly PD. An experimental trial exploring the impact of continuous transdermal alcohol monitoring upon alcohol consumption in a cohort of male students. PLoS One 2013 Jun 25;8(6):e67386 [FREE Full text] [doi: 10.1371/journal.pone.0067386] [Medline: 23825656]
25. Norman T, Peacock A, Ferguson SG, Kuntsche E, Bruno R. Combining transdermal and breath alcohol assessments, real-time drink logs and retrospective self-reports to measure alcohol consumption and intoxication across a multi-day music festival. Drug Alcohol Rev 2021 Nov 10;40(7):1112-1121. [doi: 10.1111/dar.13215] [Medline: 33174260]
26. Villalba K, Cook C, Dévieux JG, Ibanez GE, Oghogho E, Neira C, et al. Facilitators and barriers to a contingency management alcohol intervention involving a transdermal alcohol sensor. Heliyon 2020 Mar;6(3):e03612 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03612] [Medline: 32258468]
27. Dougherty DM, Hill-Kapturczak N, Liang Y, Karns TE, Cates SE, Lake SL, et al. Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use. Drug Alcohol Depend 2014 Sep 01;142:301-306. [doi: 10.1016/j.drugalcdep.2014.06.039] [Medline: 25064019]
28. Mathias CW, Hill-Kapturczak N, Karns-Wright TE, Mullen J, Roache JD, Fell JC, et al. Translating transdermal alcohol monitoring procedures for contingency management among adults recently arrested for DWI. Addict Behav 2018 Aug;83:56-63 [FREE Full text] [doi: 10.1016/j.addbeh.2018.01.033] [Medline: 29397211]
29. Dougherty DM, Karns TE, Mullen J, Liang Y, Lake SL, Roache JD, et al. Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking. Drug Alcohol Depend 2015 Mar 01;148:77-84 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.12.021] [Medline: 25582388]
30. Averill F, Brown TG, Robertson RD, Tchomgang A, Berbiche D, Nadeau L, et al. Transdermal alcohol monitoring combined with contingency management for driving while impaired offenders: a pilot randomized controlled study. Traffic Inj Prev 2018 Jul 04;19(5):455-461. [doi: 10.1080/15389588.2018.1448079] [Medline: 29543499]
31. Richards VL, Wang Y, Porges EC, Gullett JM, Leeman RF, Zhou Z, et al. Using alcohol biosensors and biomarkers to measure changes in drinking: associations between transdermal alcohol concentration, phosphatidylethanol, and self-report in a contingency management study of persons with and without HIV. Exp Clin Psychopharmacol 2023 Dec;31(6):991-997. [doi: 10.1037/pha0000637] [Medline: 36649152]
32. Ariss T, Fairbairn CE, Bosch N. Examining new-generation transdermal alcohol biosensor performance across laboratory and field contexts. Alcohol Clin Exp Res (Hoboken) 2023 Jan;47(1):50-59 [FREE Full text] [doi: 10.1111/acer.14977] [Medline: 36433786]
33. Ash GI, Gueorguieva R, Barnett NP, Wang W, Robledo DS, DeMartini KS, et al. Sensitivity, specificity, and tolerability of the BACTrack Skyn compared to other alcohol monitoring approaches among young adults in a field-based setting. Alcohol Clin Exp Res 2022 May;46(5):783-796 [FREE Full text] [doi: 10.1111/acer.14804] [Medline: 35567595]
34. Courtney JB, Russell MA, Conroy DE. Acceptability and validity of using the BACtrack skyn wrist-worn transdermal alcohol concentration sensor to capture alcohol use across 28 days under naturalistic conditions - a pilot study. Alcohol 2023 May;108:30-43 [FREE Full text] [doi: 10.1016/j.alcohol.2022.11.004] [Medline: 36473634]
35. Brobbin E, Deluca P, Coulton S, Drummond C. Accuracy of transdermal alcohol monitoring devices in a laboratory setting. Alcohol Alcohol 2024 Jan 11;59(1):agad068. [doi: 10.1093/alcalc/agad068] [Medline: 37873967]
36. Brobbin E, Parkin S, Deluca P, Drummond C. A qualitative exploration of the experiences of transdermal alcohol sensor devices amongst people in receipt of treatment for alcohol use disorder. Addict Behav Rep 2024 Jun;19:100544 [FREE Full text] [doi: 10.1016/j.abrep.2024.100544] [Medline: 38596194]
37. Didier NA, King AC, Polley EC, Fridberg DJ. Signal processing and machine learning with transdermal alcohol concentration to predict natural environment alcohol consumption. Exp Clin Psychopharmacol 2024 Apr;32(2):245-254. [doi: 10.1037/pha0000683] [Medline: 37824232]
38. Gunn RL, Merrill JE, Haines AM, Fernandez ME, Souza T, Berey BL, et al. Use of the BACtrack Skyn alcohol biosensor: practical applications for data collection and analysis. Addiction 2023 Aug;118(8):1586-1595 [FREE Full text] [doi: 10.1111/add.16207] [Medline: 37060272]
39. Kianersi S, Ludema C, Agley J, Ahn YY, Parker M, Ideker S, et al. Development and validation of a model for measuring alcohol consumption from transdermal alcohol content data among college students. Addiction 2023 Oct 30;118(10):2014-2025 [FREE Full text] [doi: 10.1111/add.16228] [Medline: 37154154]
40. Richards VL, Rajendran S, Cook RL, Leeman RF, Wang Y, Prins C, et al. Identifying desired features that would be acceptable and helpful in a wrist-worn biosensor-based alcohol intervention: interview study among adults who drink heavily. J Med Internet Res 2023 Feb 02;25:e38713 [FREE Full text] [doi: 10.2196/38713] [Medline: 36729576]

41. Richards VL, Barnett NP, Cook RL, Leeman RF, Souza T, Case S, et al. Correspondence between alcohol use measured by a wrist-worn alcohol biosensor and self-report via ecological momentary assessment over a 2-week period. Alcohol Clin Exp Res (Hoboken) 2023 Feb 21;47(2):308-318 [FREE Full text] [doi: 10.1111/acer.14995] [Medline: 36507857]
42. Wang Y, Fridberg DJ, Leeman RF, Cook RL, Porges EC. Wrist-worn alcohol biosensors: strengths, limitations, and future directions. Alcohol 2019 Dec;81:83-92 [FREE Full text] [doi: 10.1016/j.alcohol.2018.08.013] [Medline: 30179709]
43. Drug misuse in over 16s: psychosocial interventions clinical guideline. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/cg51 [accessed 2024-04-29]
44. Drake RE, O'Neal EL, Wallach MA. A systematic review of psychosocial research on psychosocial interventions for people with co-occurring severe mental and substance use disorders. J Subst Abuse Treat 2008 Jan;34(1):123-138. [doi: 10.1016/j.jsat.2007.01.011] [Medline: 17574803]
45. Dutra L, Stathopoulou G, Basden SL, Leyro TM, Powers MB, Otto MW. A meta-analytic review of psychosocial interventions for substance use disorders. Am J Psychiatry 2008 Feb;165(2):179-187. [doi: 10.1176/appi.ajp.2007.06111851] [Medline: 18198270]
46. Prendergast M, Podus D, Finney J, Greenwell L, Roll J. Contingency management for treatment of substance use disorders: a meta-analysis. Addiction 2006 Nov 10;101(11):1546-1560. [doi: 10.1111/j.1360-0443.2006.01581.x] [Medline: 17034434]
47. Petry NM. A comprehensive guide to the application of contingency management procedures in clinical settings. Drug Alcohol Depend 2000 Feb 01;58(1-2):9-25. [doi: 10.1016/s0376-8716(99)00071-x] [Medline: 10669051]
48. Petry NM, Martin B, Cooney JL, Kranzler HR. Give them prizes, and they will come: contingency management for treatment of alcohol dependence. J Consult Clin Psychol 2000 Apr;68(2):250-257. [doi: 10.1037//0022-006x.68.2.250] [Medline: 10780125]
49. Alessi SM, Petry NM. A randomized study of cellphone technology to reinforce alcohol abstinence in the natural environment. Addiction 2013 May 30;108(5):900-909 [FREE Full text] [doi: 10.1111/add.12093] [Medline: 23279560]
50. Swift R. Direct measurement of alcohol and its metabolites. Addiction 2003 Dec 30;98 Suppl 2(s2):73-80. [doi: 10.1046/j.1359-6357.2003.00605.x] [Medline: 14984244]
51. Paton A. Alcohol in the body. BMJ 2005 Jan 08;330(7482):85-87 [FREE Full text] [doi: 10.1136/bmj.330.7482.85] [Medline: 15637372]
52. Litten RZ, Bradley AM, Moss HB. Alcohol biomarkers in applied settings: recent advances and future research opportunities. Alcohol Clin Exp Res 2010 Jun 25;34(6):955-967. [doi: 10.1111/j.1530-0277.2010.01170.x] [Medline: 20374219]
53. McDonell MG, Leickly E, McPherson S, Skalisky J, Srebnik D, Angelo F, et al. A randomized controlled trial of ethyl glucuronide-based contingency management for outpatients with co-occurring alcohol use disorders and serious mental illness. Am J Psychiatry 2017 Apr 01;174(4):370-377 [FREE Full text] [doi: 10.1176/appi.ajp.2016.16050627] [Medline: 28135843]
54. Oluwoye O, Reneau H, Herron J, Alcover KC, McPherson S, Roll J, et al. Pilot study of an integrated smartphone and Breathalyzer contingency management intervention for alcohol use. J Addict Med 2020;14(3):193-198 [FREE Full text] [doi: 10.1097/ADM.0000000000000553] [Medline: 31567597]
55. Lauckner C, Taylor E, Patel D, Whitmire A. The feasibility of using smartphones and mobile breathalyzers to monitor alcohol consumption among people living with HIV/AIDS. Addict Sci Clin Pract 2019 Nov 26;14(1):43 [FREE Full text] [doi: 10.1186/s13722-019-0174-0] [Medline: 31771655]
56. You CW, Chen YC, Chen CH, Lee CH, Kuo PH, Huang MC, et al. Smartphone-based support system (SoberDiary) coupled with a Bluetooth breathalyser for treatment-seeking alcohol-dependent patients. Addict Behav 2017 Feb;65:174-178. [doi: 10.1016/j.addbeh.2016.10.017] [Medline: 27816043]
57. Hämäläinen MD, Zetterström A, Winkvist M, Söderquist M, Karlberg E, Öhagen P, et al. Real-time monitoring using a breathalyzer-based eHealth system can identify lapse/relapse patterns in alcohol use disorder patients. Alcohol Alcohol 2018 Jul 01;53(4):368-375. [doi: 10.1093/alcalc/agy011] [Medline: 29590325]
58. Goldfine C, Lai JT, Lucey E, Newcomb M, Carreiro S. Wearable and wireless mHealth technologies for substance use disorder. Curr Addict Rep 2020 Sep 01;7(3):291-300 [FREE Full text] [doi: 10.1007/s40429-020-00318-8] [Medline: 33738178]
59. Brobbin E, Deluca P, Coulton S, Parkin S, Drummond C. Comparison of transdermal alcohol concentration and self-reported alcohol consumption in people with alcohol dependence attending community alcohol treatment services. Drug Alcohol Depend 2024 Mar 01;256:111122 [FREE Full text] [doi: 10.1016/j.drugalcdep.2024.111122] [Medline: 38367536]
60. Brobbin E, Parkin S, Deluca P, Drummond C. A qualitative exploration of the experiences of transdermal alcohol sensor devices amongst alcohol service practitioners (South London, UK). Addict Res Theory 2025;33(1):36-47 [FREE Full text] [doi: 10.1080/16066359.2024.2309869] [Medline: 39906322]
61. Merrill JE, Fan P, Wray TB, Miranda Jr R. Assessment of alcohol use and consequences: comparison of data collected via timeline followback interview and daily reports. J Stud Alcohol Drugs 2020 Mar;81(2):212-219 [FREE Full text] [doi: 10.15288/jsad.2020.81.212] [Medline: 32359051]
62. Miguel AQ, Madruga CS, Cogo-Moreira H, Yamauchi R, Simões V, da Silva CJ, et al. Contingency management is effective in promoting abstinence and retention in treatment among crack cocaine users in Brazil: a randomized controlled trial. Psychol Addict Behav 2016 Aug;30(5):536-543 [FREE Full text] [doi: 10.1037/adb0000192] [Medline: 27442691]

63. Sobell L, Sobell M. Timeline follow-back. In: Litten RZ, Allen JP, editors. Measuring Alcohol Consumption: Psychosocial and Biochemical Methods. Totowa, NJ: Humana Press; 1992:41-72.
64. van Egmond K, Anderson-Luxford D, Kuntsche E, Wright C, Caluzzi G, Pennay A. Measuring alcohol consumption while watching sport events: a feasibility and validity study comparing ecological momentary assessments and transdermal alcohol monitors. Journal of Substance Use 2023 Apr 20;29(5):689-695. [doi: 10.1080/14659891.2023.2199850]
65. Dougherty DM, Moon TJ, Liang Y, Roache JD, Lamb RJ, Mathias CW, et al. Effectiveness of contingency management using transdermal alcohol monitoring to reduce heavy drinking among driving while intoxicated (DWI) arrestees: a randomized controlled trial. Alcohol Clin Exp Res (Hoboken) 2023 Oct 13;47(10):1989-2001 [FREE Full text] [doi: 10.1111/acer.15180] [Medline: 37864527]

Abbreviations

AUD: alcohol use disorder
BCa: bias–corrected and accelerated
CM: contingency management
TAC: transdermal alcohol concentration
TAS: transdermal alcohol sensor
TLFB: timeline followback

Edited by A Kushniruk; submitted 24.07.24; peer-reviewed by R Zhang, O Zhabenko; comments to author 14.10.24; revised version received 18.10.24; accepted 20.12.24; published 14.03.25.

Please cite as:
Brobbin E, Drummond C, Parkin S, Deluca P
Use of Wearable Transdermal Alcohol Sensors for Monitoring Alcohol Consumption After Detoxification With Contingency Management: Pilot Randomized Feasibility Trial
JMIR Hum Factors 2025;12:e64664
URL: https://humanfactors.jmir.org/2025/1/e64664
doi:10.2196/64664
PMID:

Patient Experiences With a Mobile Self-Care Solution for Low-Complex Orthopedic Injuries: Mixed Methods Study

Jelle Spierings[1*], MD; Gijs Willinge[2*], MD; Marike Kokke[1*], MD; Sjoerd Repping[3*], PhD; Wendela de Lange[4], BSc; Thijs Geerdink[2*], MD, PhD; Ruben van Veen[2*], MD, PhD; Detlef van der Velde[1*], MD, PhD; Carel Goslings[2*], MD, PhD; Bas Twigt[2*], MD, PhD; Collaboration Group[6]

[1]Department of Trauma Surgery, St. Antonius Ziekenhuis Utrecht, Koekoekslaan 1, Nieuwegein, The Netherlands

[2]Department of Trauma Surgery, OLVG Hospital, Amsterdam, The Netherlands

[3]University of Amsterdam, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[4]The Healthcare Innovation Center, Julius Centrum, University Medical Center Utrecht, Utrecht, The Netherlands

6

[*]these authors contributed equally

Corresponding Author:
Jelle Spierings, MD
Department of Trauma Surgery, St. Antonius Ziekenhuis Utrecht, Koekoekslaan 1, Nieuwegein, The Netherlands

Abstract

Background: The Dutch acute health care system faces challenges with limited resources and increasing patient numbers. To reduce outpatient follow-up, direct discharge (DD) has been implemented in over 30 out of 80 Dutch hospitals. With DD, no routine follow-up appointments are scheduled after the emergency department (ED) visit for low-complex, isolated, and stable musculoskeletal injuries. This policy is supported by information leaflets, a smartphone app, and a telephone helpline with human support. Growing evidence shows that DD is satisfactory, safe, and effective in reducing secondary health care use, but thorough patient experiences are lacking.

Objective: The aim of this study was to explore the experiences of patients with DD to ensure durable adoption and to improve the treatment protocol.

Methods: A mixed method study was conducted parallel to the implementation of DD in 3 hospitals. Data were collected through a survey directly after the ED visit, a survey 3 months post injury, and semistructured interviews. Quantitative data were reported descriptively, and qualitative data used thematic analysis. Outcomes included the Bowen feasibility parameters: implementation, acceptance, preliminary efficacy, and demand. All patients who consented to the study face-to-face with one of the 12 low-complex musculoskeletal injuries were included in the study during the implementation period.

Results: Of the 429 patients who started the primary survey, 138 patients completed both surveys. A total of 18 semistructured interviews were conducted and analyzed. Patients reported a median treatment satisfaction score of 7.8 (IQR 6.6-8.8) on a 10-point scale of DD at the ED. Information quality was experienced as good (106/138, 77%), and most preferred DD over face-to-face follow-up (79/138, 59%). Patient information demands and app use varied among patients, with a median frequency of use of 4 times (ranging from 1 to 30).

Conclusions: This study shows that patients consider DD a feasible and safe alternative to traditional treatment, with a favorable perception of its acceptability, efficacy, applicability, and demand. Nevertheless, response rates were relatively low, and personal nuances and preferences must be considered when implementing DD. Clinicians and policy makers can use the insights to improve DD and work towards the integration of DD into clinical practice and future guidelines.

(JMIR Hum Factors 2025;12:e53074) doi:10.2196/53074

KEYWORDS

self-care application; mHealth; experience; traumasurgery; orthopedic surgery; virtual fracture clinic; patient perspective; direct discharge; musculoskeletal injury; mobile self-care; method study; health care system; hospital; mobile health; app; smartphone; satisfactory; effectiveness; treatment; virtual clinic; virtual care; digital health

Introduction

The Dutch acute health care system faces substantial challenges due to limited resources and a rising number of patients requiring in-hospital care [1,2]. To alleviate this pressure, virtual fracture clinics have been introduced for Orthopedic and Trauma surgery patients [3]. Direct discharge (DD) is the most basic part of a virtual fracture clinic, concerning solely low-complex, isolated, stable musculoskeletal injuries. With DD, patients with these injuries are discharged directly from the emergency department (ED) without routine outpatient follow-up. Patients receive a removable orthosis or sling and are given information summarized in a mobile self-care app, the Virtual Fracture Care (VFC) app.

Growing evidence shows that DD is a safe and effective alternative to "traditional" care with routine follow-up [4,5]. DD reduces secondary health care use without causing a shift to primary health care use, while patient-reported outcomes (eg, functional outcome and satisfaction scores) and adverse outcomes (eg, complications and persistent complaints) are comparable [3,4]. These results and the usefulness of this method during COVID-19 social distancing measurements have led to a rapid uptake of DD in the Netherlands for 12 frequent injuries at the ED [6]. Since the first introduction in 2019, over 25 out of the 80 Dutch hospitals have implemented DD as the standard of care, adding to over 100 virtual fracture clinics and DDs worldwide [7]. While this reorganization of care has yielded beneficial outcomes in terms of reducing resources with comparable patient outcomes, there is a lack of research on the experiences of patients and their relatives with DD.

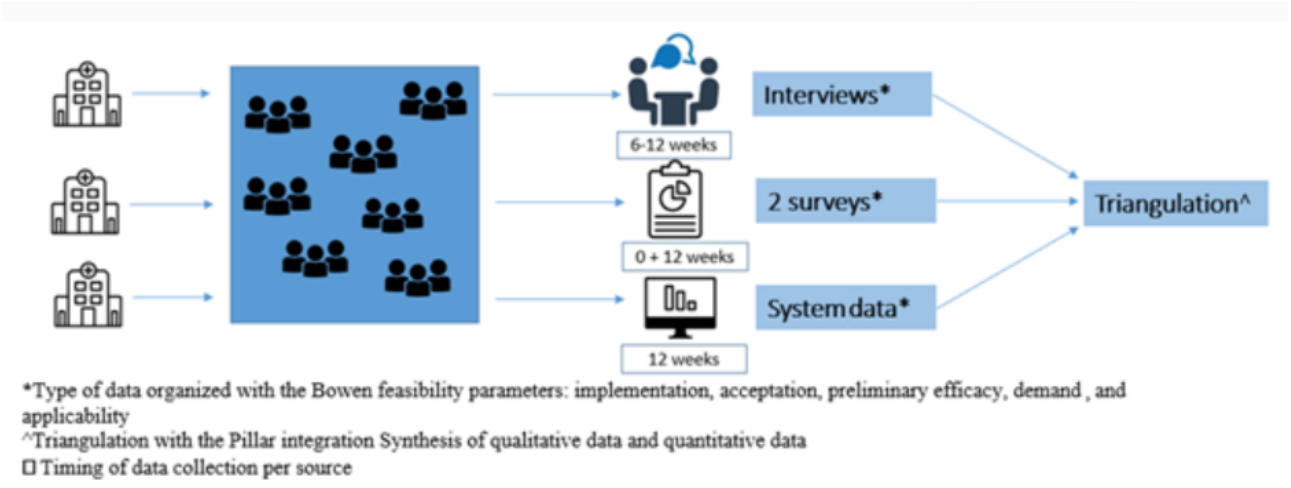
Performing an in-depth evaluation from an end user perspective deepens the insight into the quality of care by gaining a deeper understanding of user experiences and exploring reasons for nonadoption or abandonment, ensuring sustainable adoption of DD (inter)nationally. The aim of this study was therefore to explore the experiences of patients with DD to ensure durable adoption and to improve the protocol.

Methods

Design

An observational mixed method study was conducted among patients and parents of patients younger than 12 years who sustained low-complex, isolated, stable musculoskeletal injuries parallel to the implementation of DD in 3 Dutch level-2 trauma centers from September 2021 to July 2022 with an inclusion period of 3 months per hospital (Figure 1). Quantitative data and qualitative data were collected and analyzed separately by a quantitative team (GW and JS) and a qualitative team (WL and E Mathijsen). The Bowen feasibility framework was used to organize the data within the following parameters: implementation, acceptation, preliminary efficacy, and demand [8]. After separate analyses, quantitative data and qualitative data were triangulated with the Pillar Integration Process [9]. This study was reported according to the Good Reporting on a Mixed Methods Study (GRAMMS) criteria (Multimedia Appendix 1) [10], and according to the "Improving and Standardizing Evaluation Reports of Web-based and Mobile Health Interventions from the CONSORT-EHEALTH" [11].

Figure 1. Summary of study procedures and models used to evaluate the direct discharge protocol among patients.



Context

All participating centers were urban, level-2 trauma centers and teaching hospitals with up to 3 locations per hospital. All locations per hospital have an ED and implemented DD at the same time. Each hospital treats between 1200 and 1800 patients with low-complex, stable, isolated musculoskeletal injuries annually.

Traditional Treatment

Before DD was implemented, patients were treated according to local trauma protocols. These protocols consisted of immobilization or support with either a cast, sling, bandage, or splint and brief information about the injury at the ED. At least 1 outpatient follow-up appointment was scheduled at the plaster room or in the outpatient clinic within 2 weeks after the injury for review, extensive information, and definitive management planning.

Direct Discharge Protocol

This protocol was derived from the British model of a virtual fracture clinic and adapted to the Dutch health care setting in 2018. In its Dutch adaptation, DD includes 12 treatment protocols for low-complex, stable traumatic orthopedic injuries with additional injury-related criteria (Multimedia Appendix 2) [6]. Patients who met the injury-related inclusion criteria in Multimedia Appendix 2 and spoke Dutch or English fluently were included. No further predefined restrictions (eg, age or comorbidity) were used. Patients were excluded from the protocol if they had initial treatment in another hospital, follow-up in another hospital (eg, closer to home), multiple injuries, the reason for follow-up other than the injury (eg, social-care reasons), Eye/Motor/Verbal-score<15 at presentation, or intoxication. With DD, patients were discharged directly from the ED without routine outpatient follow-up. They receive a removable orthosis or a sling (eg, brace instead of a cast), and extensive information at the ED, summarized in a mobile self-care app, the VFC app. Apart from the criteria in Multimedia Appendix 2 and Dutch or English language skills (level B1), the protocols did not contain any predefined restrictions (eg, age or comorbidity). If, however, the ED staff believed physical follow-up was the most suitable treatment, outpatient follow-up was scheduled accordingly.

Patient eligibility for the protocol was re-evaluated on the next workday (within 24 h) by a team consisting of an (orthopedic) trauma surgeon and a radiologist. Patients who were incorrectly discharged directly were contacted by phone and scheduled for a face-to-face appointment.

VFC App

The VFC app provides self-care assistance through information, videos, and a helpline and can be downloaded for free at the Google Play Store and iOS App Store (Figure 2). Injury-specific leaflets with recovery information, treatment rules, and red flags were included. Furthermore, frequently asked questions, audiovisual exercise-, immobilization-, and analgesic instructions were included to assist patients. If patients required human contact in addition to the information, a helpline operated by an employee (eg, plaster technician) was available during working hours. The VFC app aimed to increase self-management and self-care during recovery and to substitute face-to-face follow-up. No reminders were sent, and use was voluntary. The app was developed by OLVG to reduce health care use and built by medical doctors with IT experience, usability testing was performed with peers. Due to its success, it was implemented in other hospitals. No major changes occurred during the study period.

Figure 2. English and Dutch in-app screenshots of the Virtual Fracture Care app used in the direct discharge protocol.

Study Population

Patients who met inclusion criteria and consented to participate in the surveys in the VFC app were included. Based on the annual number of patients (1200 - 1800 per hospital), and the inclusion time of 3 months per hospital, the estimated number of eligible patients was between 900 and 1350. Parents were asked to fill out the surveys if patients were 12 years or younger. If patients were between 13 and 16 years of age, patients and parents were allowed to complete the surveys. Patients older than 16 years could fill out the surveys alone. The exclusion criteria for this study were the same as the previously mentioned exclusion criteria for DD.

Sampling and Recruitment

Before discharge from the ED, in a face-to-face setting, eligible patients were asked to download and open the VFC app. An in-app pop-up asked for informed consent to participate in 2 surveys and an interview. After informed consent, patients were redirected to a Research Electronic Data Capture (REDCap) environment, a web-based survey system, to fill out the primary survey and were given an opt-out form by the treating physician, [12]. Additional information was given about the study and withdrawal methods. In the REDCap environment, all participants who started the survey were invited to participate in the interview by clicking a button and providing their email addresses. Age, sex, type of injury, and hospital were used to select a purposive sample among the quantitative population for interview patients. Eligible patients were contacted a

minimum of 6 weeks after injury to schedule a semistructured interview. During data collection, authors WDL and E Mathijsen considered whether the qualitative data had gained sufficient depth to perform a thorough analysis. Patients were reminded by email to complete the survey, and the second survey was sent 3 months post injury. Additional patient data were collected from electronic patient records.

Data Collection

Data were collected from surveys, semistructured interviews, and system data.

Surveys

Two surveys, 1 directly after the ED visit and one 3 months after the ED visit, with 63 questions, including close-ended questions, multiple-choice questions, 5-point Likert Scales, visual analog scales, and free-text questions, measured 5 Bowen feasibility parameters (Multimedia Appendix 3) [8]. As no golden standard for the evaluation of innovations exists, the surveys and topic list were developed by 4 researchers (JS, GW, BT, and THG) and checked by 2 experts on relevance: a professor in trauma surgery (JC Goslings) and an associate professor in-process evaluations of health care innovations (JCA Trappenburg). We pretested the survey with 5 patients to improve clarity.

Semistructured Interviews

Two independent researchers specialized in qualitative research, not involved in daily clinical practice or the VFC research team, conducted digital semistructured interviews to minimize social-desirability bias. The interviews were held within 6 and 10 weeks post injury to warrant an optimal recall. The interviews were guided by a topic list based on literature, including previously mentioned Bowen feasibility parameters (Multimedia Appendix 4) [8]. The research team piloted the topic list for clarity and completeness and modified it during data collection.

System Data

Quantitative data were extracted from the electronic patient record after 3 months of follow-up. The patient characteristics included compliance to therapy, complications (yes/no), type of complications, follow-up (yes/no), type of follow-up, and imaging. Downtime from the app was extracted from the log record of the VFC app.

Data Analysis

Quantitative data were analyzed using the SPSS (version 27; IBM Corporation) [13]. Baseline characteristics were reported descriptively using numbers and proportions for categorical variables and mean with SD or median with IQR as appropriate. The normal distribution of continuous data was assessed with visual analysis. Discrete variables were reported as numbers (percentages of the total population). The paired *t* test or the Mann-Whitney *U* test was used to determine the statistical significance of parametric variables for normally and nonnormally distributed data.

Qualitative data were analyzed according to the principles of the 6 phases of thematic analysis by Braun and Clark [14]. We have used an inductive, categorical approach because the triangulation process started after 4 phases. Data analysis started after the first 5 interviews. Interviews were audiotaped, transcribed verbatim, and analyzed using the software program NVivo (QSR International) [14]. One researcher (WL) independently analyzed the data, and another researcher (E Mathijsen) reviewed the analysis. Two researchers (WL and E Mathijsen) used several methods to ensure reliability and validity [15-17]. Discrepancies and remarks were discussed until they reached a consensus about data interpretations. Memos were made to track research decisions during analysis. Code saturation was reached when no new categories or themes emerged from the new raw data [15,16]. Instead of steps 5 and 6 according to Braun and Clark [14], we organized the data per theme during the triangulation session. The final themes were used to describe the value and feasibility of DD from the perspective of patients.

Triangulation

After the separate quantitative and partly qualitative analyses, findings were triangulated with a simplified approach of the Pillar Integration Process technique [9]. This approach uses a transparent and rigorous 4-stage technique for integrating and presenting qualitative and quantitative findings in a joint display Microsoft Excel, 2018 [18]. One of the researchers presented the quantitative findings (JS) per study parameter, and another the qualitative findings (WDL). (Dis)similarities and self-contained themes were objectified. One of the researchers (E Mathijsen) merged these themes into a meaningful narrative (the pillar), reviewed by researchers JS and WDL.

Ethical Considerations

This study, including the process analysis, was reviewed and approved by the Medical Ethical Committee of Utrecht, Netherlands (W21.261). Patients provided consent for participation in the research and could opt out at any time after request by email. The original consent and institutional review board approval covers secondary analysis without additional consent. A data key is stored at the local hospitals in a secured map and coded file. This is only accessible to JS and GW. The accessible data have been deidentified as far as possible (eg, age in years instead of the date of birth). Patients received no compensation to participate in this research.

Results

Demographics

Of the 429 patients who started the primary survey, 203 did not provide any data or contact details and did not complete the first survey and 88 did not complete both surveys (Figure 3). Of the 138 patients that completed both surveys (response rate: 138/429, 32%), 83 out of 138 (60%) were female, and the median age was 50 years (IQR 12 to 61) (Table 1). Patients who provided contact details at baseline varied significantly from responders regarding age (P=.01) but not sex (P=.14). Most patients were native Dutch speakers, who had attended primary school in the Netherlands (136/198, 98%), and over half had a minimum of a bachelor's degree (82/198, 59%) (Table 1). In addition, 18 patients sampled from the quantitative source participated in a web-based semistructured interview, of which

15 (83%) patients were female and 6 (33%) patients were parents of children (Table 2).

Figure 3. Flow diagram of eligible patients to evaluate the DD protocol. DD: direct discharge.

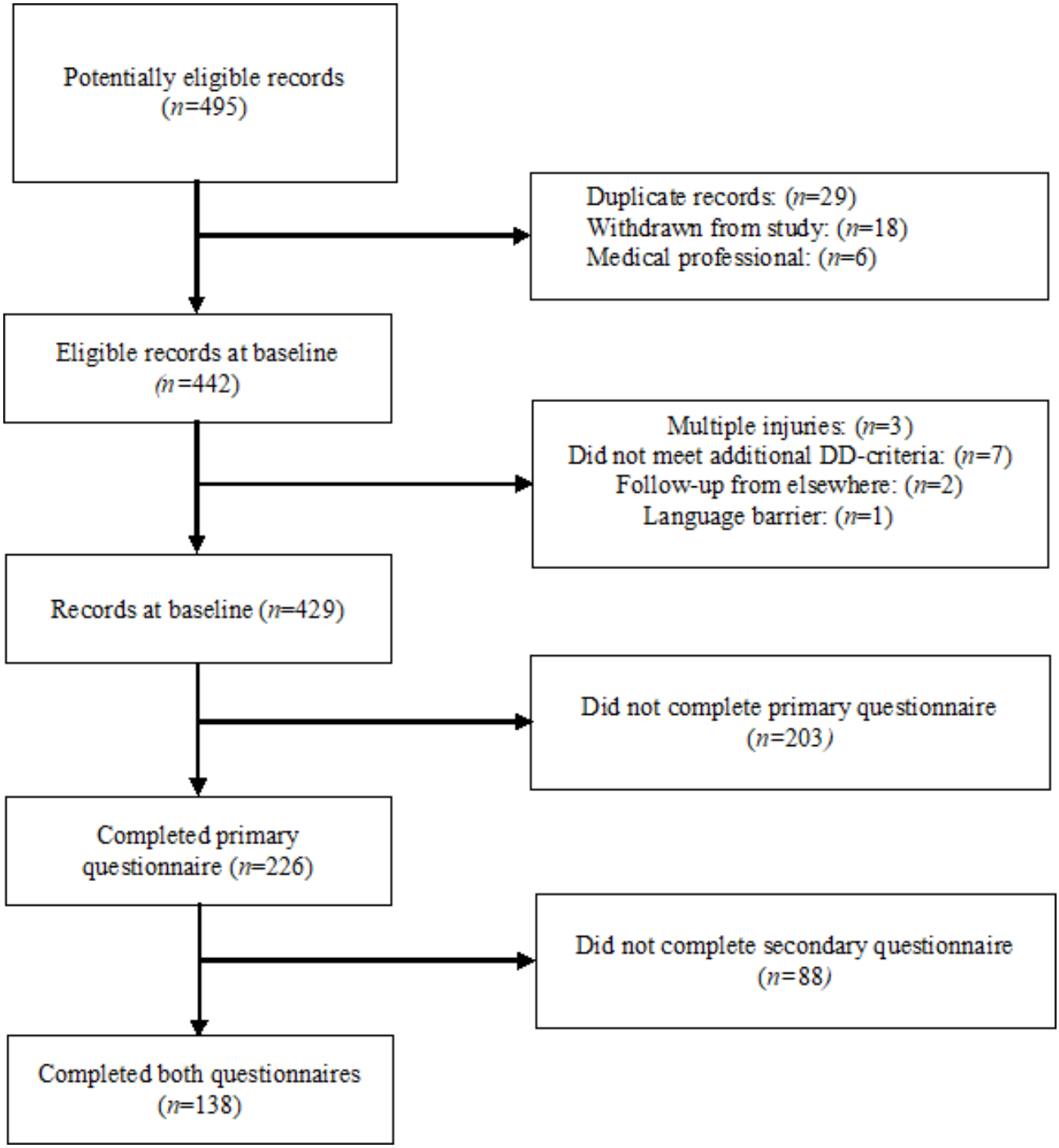


Table 1. Baseline characteristics of included patients in quantitative analysis of direct discharge.

Characteristic	Included in quantitative analysis (n=138)
Age (years), median (IQR)	50 (12 - 61)
Age younger than 18 years, n (%)	51 (37)
Sex (female), n (%)	83 (60)
Country of primary school, n (%)	
Netherlands	136 (98)
Philippines	1 (1)
Italy	1 (1)
Primary language, n (%)	
Dutch	136 (98)
Filipino	1 (1)
Italian	1 (1)
Highest level of education, n (%)	
Elementary school	4 (3)
High school	21 (15)
Vocational school	28 (20)
College	50 (36)
University	32 (23)
Other	3 (2)
Hospital treated, n (%)	
Center A	54 (39)
Center B	26 (19)
Center C	58 (42)

Table 2. Demographics of interviewed patients or parents regarding experiences with the direct discharge protocol.

ID	Sex	Type of injuries	Year of birth
#1	Female	Mallet finger	1969
#2	Female	Weber A Fx[a] or avulsion Fx	1971
#3	Female	Weber A Fx or avulsion Fx	2000
#4	Male	Fifth metatarsal Fx	1956
#5	Female	Radial head or neck Fx	1957
#6	Female	Ankle distortion	1992
#7	Female	Radial head or neck Fx	1963
#8	Female	Torus Fx of the radius	2011
#9	Male	Radial head or neck Fx	1959
#10	Female	Hallux Fx	1964
#11	Male	Fifth metatarsal Fx	1980
#12	Female	Hallux Fx	2010
#13	Female	Spoke injury	2017
#14	Female	Greenstick Fx	2014
#15	Female	Radial head or neck Fx	1974
#16	Female	Torus Fx	2014
#17	Female	Torus Fx	2011
#18	Female	Weber A Fx or avulsion Fx	1956

[a]Fx: fracture.

Implementation

Both data sources indicated that most patients were satisfied with their ED visit and the introduction of the app, despite some mentioning the hectic ED environment and difficulty downloading the app due to poor Wi-Fi connection. The physiological distress, and the hectic ED environment, resulted in an inability to recollect all the information provided by the ED physician. Several patients pointed out that the information in the VFC app was a valuable supplement to their ED visit (quote 1). Mobile app stability was excellent, with only 1 patient reporting issues accessing the app. The helpline could not be reached in 4 cases, resulting in 2 patients contacting the ED directly. The app was available 99.6% of the time, with the only downtime (5 hours in 3 months) caused by a software problem that was fixed by the app builder.

> *Because… it always goes quickly. It is always busy. But still, you know, this was completely clear; what I could expect and that I could download the app.* [quote 1; ID #10]

Acceptation

The median satisfaction with treatment was 7.8 (IQR 6.6-8.8) on a 10-point scale. Qualitative data complemented quantitative data, as most patients mentioned that they were satisfied with DD (quotes 2 and 3).

> *I would give DD a 7 or 8. Yeah, let's say 7, because I do feel that it might be difficult for older people. Especially because they don't always understand technology, you know.* [quote 2; ID #15]

> *To be honest, I think it's better. You know, often it's like, you go to a hospital, and it takes half a day just to get there, and come back, and all those things, and then it's just like: "O, it looks fine."* [quote 3; ID #5]

VFC App Acceptation

Most patients (106/138, 77%) reported that the quality of the information in the app was good and reported that the information answered the questions they had during their recovery (82/138, 58%). Additionally, 59% (79/138) would prefer the app over face-to-face follow-up if they were to have a similar injury. However, in qualitative data, some patients expressed concerns that the app may not be suitable for people with limited digital skills. Even though none of the patients identified themselves as "vulnerable," many reported that the app may not be appropriate for vulnerable individuals. Patients expressed that DD should not be mandatory for all in order to protect potentially vulnerable patients for whom DD would not be suitable.

Brace Acceptation

A total of 122 patients out of 138 (89%) used the brace for the prescribed period. Most patients (116/138, 84%) removed the brace or the sling during showering, at night (118/138, 86%), and (72/138, 52%) during non–weight bearing exercises. Most patients (97/138, 71%) found the brace comfortable, and 94 out of 138 patients (68%) found the brace convenient for these types of injuries (Multimedia Appendix 5). These findings were

complementary to the qualitative data. Patients preferred the brace to a cast as it allowed for better mobility because it was less bulky and less rigid compared with a cast (quote 4). However, the increased mobility made it tempting to overexert oneself, resulting in increased pain and insecurity for some patients (quote 5). The smaller size of the braces made some injuries seem less serious and burdensome, reportedly leading to social pressure to return to work earlier than advised by the doctor or app. Additionally, the less rigid nature of the brace resulted in increased skin friction, causing a superficial ulcer in 1 patient in the qualitative data.

> *I really like the brace because it allowed me to move. So, I was not stuck with a bulky cast on my leg, but rather, I had a lot of flexibility. I could actually determine what I wanted to do or not. So, I found that very enjoyable.* [quote 4; ID #2]

> *The brace was sometimes the reason I went over my limit. I could move more and was in the process of moving to a new home. But then the pain came back. I found it very difficult to 'guard' my limits.* [quote 5; ID #18]

Helpline and Perceived Safety

Most patients found the helpline important (116/138, 84%) and that the helpline offered them a sense of security (86/138, 62%). For some patients, the helpline was a way of checking if their recovery was still on track (quote 6). Most patients (94/138, 68%) in both data sources expressed that treatment with a brace and a helpline is safe for these types of injuries. However, some expected that if injuries were more severe, they would need more assistance than a brace and a smartphone app (quote 7).

Self-Empowerment

An increase in perceived self-empowerment was reported in 67 out of 138 (49%) of all patients, 51 (37%) patients reported neutral results, and 20 (14%) patients reported no increase. Enhanced treatment engagement was reported in 58 (42%) patients, while 50 (36%) patients reported neutral results, and 31 out of 138 (23%) patients reported no enhanced engagement (Multimedia Appendix 6). In the qualitative data, some of the interviewed patients stated that this type of treatment provided more self-empowerment and therefore, more control in their recovery (quote 8).

> *After about a week or three, I had a setback. I couldn't find this in the app, so I decided to call the helpline. The pain came back, and I was afraid I had broken something or something like that. They reassured me that it could not happen so quickly and that I just needed to rest for 24 hours h. They were right! It was nice to be able to check this.* [quote 6; ID #11]

> *Yes, I believe it is safe, given the circumstances. Because it wasn't that serious. Yes. I do have the confidence that if it is something serious I won't receive this type of treatment.* [quote 7; ID #17]

> *It was nice that I could read what I was allowed to do and what I was not. I think that gave me more control over my recovery. I knew what I could do myself in terms of exercises, and that was very helpful.* [quote 8; ID #2]

Preliminary Efficacy

Secondary Health Care Use

A total of 10 out of 138 (7%) patients had a face-to-face follow-up, and 9 (7%) patients by phone. Two patients attended the ED again after discharge due to anxiety and pain at the fracture site and were scheduled for outpatient follow-up. Two patients received follow-up for a wound check and 2 received follow-up as decided by the medical specialist due to severe pain complaints at the ED.

Functional Outcome

More than half of the patients had fully recovered 3 months after injury in terms of daily activities, sports, and work (84/138, 61%). Between 31% and 40% of patients were limited in functional outcomes once or twice a week. Approximately one in 10 patients were limited in physical activities more than 3 times a week. (Multimedia Appendix 7).

Pain

Most (92/138, 67%) patients have used painkillers in the first 3 weeks of recovery. Of these patients, 70 out of 92 (76%) patients have used acetaminophen, 17 (12%) patients have used nonsteroidal anti-inflammatory drugs, and 5 (4%) patients have used other undefined analgesics. Few patients (7/92, 5%) used cooling of the injury site to reduce pain. Analgesic use in the second week was lower (35/92, 38%) and further declined in the third week (9/92, 10%) and in the fourth week or after (9/92, 10%). Qualitative data showed limited pain complaints after immobilization of the initial trauma. Pain complaints in the first weeks were treated with analgesics. After the immobilization, a few patients had persistent pain symptoms. Most of these patients sustained a Weber A ankle fracture or ankle distortion. Those patients were most limited in their daily activities, specifically during more intense physical activities (eg, labor and sports) (quote 9).

> *The eight weeks of recovery are over, I believe. So now I should be able to start exercising again, well I do karate but I'll wait a little longer for that.* [quote 9; ID #4]

Demand

Almost all patients used the app during recovery (133/138, 95%) with a median of 4 times (IQR 2-6.5, range 0 - 30). Reasons to use the app included checking recovery exercises (91/138, 68%), treatment rules (69/133, 51%), the phase of recovery (68/133, 51%), seeking helpline information (57/133, 42%), and analgesics (10/133, 8%). Qualitative data supported the quantitative data. App use was focused on the first weeks post injury, and parents consulted the app more than children. They occasionally showed it to their child, predominantly if the information contained photos and videos (quote 10). Among the interviewed patients, almost all patients reported not requiring face-to-face follow-up for these types of injuries. The biggest advantage of DD is that it is time-saving on a personal level. Nevertheless, a few patients expressed a preference for outpatient follow-up. Especially if they had persistent complaints

or when the recovery was slower than expected. Patients expressed that this led to insecurity, which was also expressed by patients who used the app more frequently or later in the recovery phase (quote 11).

A minority of patients reported a lack of human contact and the physicians' reassurance of adequate recovery. Subsequently, as a minimal substitute, these patients suggested a feedback system (eg, pain scores or communication tools) to assist them further.

> *As for the app, he did see it, but you know, whatever! The only thing they find really interesting are pictures. What I liked that in the app is that they actually showed what was going on and what it looks like.* [quote 10; ID #17]

> *I was constantly looking for confirmation in the app, online, or at the fracture helpline.* [quote 11; ID #11]

Discussion

Principal Findings

This study shows that patients consider DD a feasible and safe alternative to traditional treatment, with a favorable perception of its acceptability, efficacy, applicability, and demand. Nevertheless, response rates were relatively low, and personal nuances and preferences must be considered when implementing DD.

Comparison With Literature

Patient satisfaction and perceived safety with DD aligned with previous studies [3,19,20]. Most patients responded positively to the introduction of DD at the ED, brace treatment, and assistance with the VFC app and helpline. The VFC app proved valuable in overcoming low recall of verbal information due to the chaotic ED environment and psychological distress. This finding is consistent with a systematic review highlighting the benefits of additional visual and written information in enhancing recall and injury knowledge [21]. Adequate understanding of the injury is crucial for properly following self-care protocols and monitoring red flags, this aligns with the finding that almost all patients used the app at least once. The braces used with DD were well-received, offering advantages such as easy removal (eg, during or at night), and improved daily living activities due to the early weight bearing. Braces seem to affect patients positively, as the lack of these advantages has been reported as most burdensome during cast immobilization [22]. However, the perceived decreased severity of the injury may pose a risk of overexertion and require further research.

Preliminary efficacy results align with previous research, demonstrating low complication rates, secondary health care use, comparable functional outcomes, and patient satisfaction with treatment [4,5,23]. Most patients preferred the VFC app over face-to-face follow-up for these injuries. While using apps in orthopedic and trauma surgery is not new, adding a self-care app to DD and virtual fracture clinics is a novel approach [24-26]. Patients found the information quality good and appreciated the time-saving component (eg, reducing work absenteeism). However, in the case of more severe injuries with wounds or complex fractures, some patients would prefer face-to-face interaction for additional reassurance. As previously suggested, digital self-care could be combined with face-to-face follow-up or used for preappointment education in patients with more severe injuries [27,28].

Although a good fit for most, some (young) patients noted that the app might not be suitable for older individuals or those with limited digital skills, potentially increasing health inequities in an already digitally oriented world [29,30]. However, it is important to note that DD targets relatively young patients, who are considered capable by the treating physician.

While previous studies have highlighted the potential for increased self-empowerment and patient engagement with eHealth, this study did not explicitly confirm it [31-33]. Despite positive study results, individual nuances in patients and injury types require a continuous assessment to ensure personalized patient care.

Information demand and app use varied among patients, implying different levels of demand among users. A minority of patients expressed a desire for more human contact and reassurance. A possible suggested solution was developing an in-app numerical feedback system, such as a questionnaire or communication tool.

Strengths and Limitations

This study has several strengths. To our knowledge, it is the first to provide in-depth interviews with patients treated using DD, providing valuable insights. Second, the research team's multidisciplinary involvement ensured a comprehensive evaluation and analysis of the data from multiple perspectives. Thirdly, the mixed method approach, including the separate collection of both data sources until data saturation, combined with the triangulation, increased the likelihood of realistic and rigorous results. Additionally, using a validated framework provided a structured insight into each feasibility parameter.

However, limitations include potential responder and selection bias due to a younger sample of patients with higher education levels and Dutch literacy rates than the general Dutch population. This age difference could be caused by the 12 selected injuries, of which half only occur in pediatric patients. Furthermore, the response rate of this study was 32%. Response rate, literacy, and education may limit the generalizability of the findings to a broader Dutch population, especially those with lower health literacy or digital skills. Furthermore, the timing to measure functional outcomes might have been suboptimal in these patients, as for some injuries (eg, mallet finger injuries), the immobilization period had just ended, resulting in limitations in daily activities.

Implications for Clinicians and Policy Makers, and Future Perspectives

DD has emerged as a promising approach to reduce outpatient follow-up while maintaining positive effects on primary health care use, patient satisfaction, complications, and functional outcomes [4,5]. By delivering follow-up care close to home, associated health care costs and societal costs decrease [34]. In addition, it limits unnecessary patient travel to the hospital thereby reducing the environmental impact of health care.

Including patients' perspectives in evaluating new care pathways, whether digitally assisted or not, is crucial for sustainable adoption and quality of care. Clinicians, researchers, and policy makers should prioritize patient involvement throughout the design, prototype, pilot, and evaluation phases. "Design thinking," a validated approach widely recognized in user experience and implementation literature, can be used to design these pathways. For example, replacing hospital care with DD involves changes in location and care deliverers, which presents new needs, challenges, and opportunities suitable to solve with design thinking. This study has identified areas for improvement of DD in terms of functions and features, and adjusted language to lower literacy. Future studies should focus on co-designing in-app feedback systems that address patient and health care professional needs for reassurance and monitoring like communication tools or questionnaires.

Conclusion

Patients consider DD a feasible and safe alternative to traditional treatment, with a favorable perception of its acceptability, efficacy, applicability, and demand. Nevertheless, response rates were relatively low, and personal nuances and preferences must be considered when implementing DD. Clinicians and policy makers can use the insights to improve DD and work towards the integration of DD into clinical practice and future guidelines.

Acknowledgments

The authors wish to thank Dr JCA Trappenburg, associate professor at the Julius Center from the University Medical Center of Utrecht, and Drs WL and E Mathijsen from the Healthcare Innovation Center (THINC) for the assistance in the development of the surveys and collecting and analyzing of the qualitative data. The institutions of the authors affiliated with St Antonius Hospital and OLVG Hospital have received funding to support this work through an unrestricted grant from the Dutch organization for health care sciences and health care innovation, ZonMw, the Hague, the Netherlands (grant 516012524).

The Colloboration Group is: Dr N Sosef, Dr T L Groenesteege, and Dr J G Ten Brinke.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Good Reporting on a Mixed Methods Study (GRAMMS) criteria for the study: "Patient Experiences With a Mobile Self-care Solution for Low-complex Orthopedic Injuries: Mixed Methods study".
[DOCX File, 26 KB - humanfactors_v12i1e53074_app1.docx]

Multimedia Appendix 2
Additional criteria and immobilization for treatment of low-complex, traumatic orthopedic injuries with the direct discharge protocol.
[DOCX File, 13 KB - humanfactors_v12i1e53074_app2.docx]

Multimedia Appendix 3
Surveys used to evaluate direct discharge among health care professionals.
[DOCX File, 163 KB - humanfactors_v12i1e53074_app3.docx]

Multimedia Appendix 4
Topic list for health care professionals the direct discharge protocol.
[DOCX File, 15 KB - humanfactors_v12i1e53074_app4.docx]

Multimedia Appendix 5
5-point Likert scale distribution of acceptance-related outcomes regarding experiences with a brace.
[DOCX File, 129 KB - humanfactors_v12i1e53074_app5.docx]

Multimedia Appendix 6
Self-empowerment and perceived safety of care with direct discharge.
[DOCX File, 73 KB - humanfactors_v12i1e53074_app6.docx]

Multimedia Appendix 7
Frequency of limitation per week in physical function, activities of daily living, and school or work in patients treated with direct discharge at 3 months follow-up.

[DOCX File, 15 KB - humanfactors_v12i1e53074_app7.docx]

References

1. Goslings JC, Gorzeman MP, M. Offeringa-Klooster JB. Brandbrief 'Regionale spoedzorg–de rek is er uit [Report in Dutch]. : TraumaNet AMC; 2016.
2. van der Linden C, Reijnen R, Derlet RW, et al. Emergency department crowding in the Netherlands: managers' experiences. Int J Emerg Med 2013 Oct 24;6(1):41. [doi: 10.1186/1865-1380-6-41] [Medline: 24156298]
3. Murphy EP, Fenelon C, Murphy RP, et al. Are virtual fracture clinics during the COVID-19 pandemic a potential alternative for delivering fracture care? A systematic review. Clin Orthop Relat Res 2020 Nov;478(11):2610-2621. [doi: 10.1097/CORR.0000000000001388] [Medline: 32657810]
4. Geerdink TH, Verbist J, van Dongen JM, Haverlag R, van Veen RN, Goslings JC. Direct discharge of patients with simple stable musculoskeletal injuries as an alternative to routine follow-up: a systematic review of the current literature. Eur J Trauma Emerg Surg 2022 Aug;48(4):2589-2605. [doi: 10.1007/s00068-021-01784-z] [Medline: 34529086]
5. Khan SA, Asokan A, Handford C, Logan P, Moores T. How useful are virtual fracture clinics? Bone Joint Open 2020 Nov 1;1(11):683-690. [doi: 10.1302/2633-1462.111.BJO-2020-0107.R1]
6. Geerdink TH, Haverlag R, van Veen RN, Bouwmeester OVA, Goslings JC. Direct discharge from the ED for patients with simple stable injuries: a Dutch pilot study. Ned Tijdschr Geneeskd 2020 Feb 25;164:D4604. [Medline: 32391996]
7. Fracture Clinic Redesign Programme. URL: http://www.fractureclinicredesign.org/visit/ [accessed 2023-03-01]
8. Bowen DJ, Kreuter M, Spring B, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457. [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
9. Johnson RE, Grove AL, Clarke A. Pillar integration process: a joint display technique to integrate data in mixed methods research. J Mix Methods Res 2019 Jul;13(3):301-320. [doi: 10.1177/1558689817743108]
10. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
11. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]
12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
13. IBM SPSS Statistics. IBM. URL: https://www.ibm.com/products/spss-statistics [accessed 2025-03-06]
14. NVivo 14 Windows. URL: https://help-nv.qsrinternational.com/14/win/Content/welcome.htm [accessed 2025-02-25]
15. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
16. Morse JM, Barrett M, Mayan M, Olson K, Spiers J. Verification strategies for establishing reliability and validity in qualitative research. Int J Qual Methods 2002 Jun;1(2):13-22. [doi: 10.1177/160940690200100202]
17. Boeije H. Analysis in Qualitative Research: Sage Publishing LTD; 2009.
18. Microsoft Excel. Microsoft. URL: https://office.microsoft.com/excel [accessed 2025-02-25]
19. Geerdink TH, Spierings JF, Schuijt HJ, et al. Healthcare utilization and satisfaction with treatment before and after direct discharge from the Emergency Department of simple stable musculoskeletal injuries in the Netherlands. Eur J Trauma Emerg Surg 2022 Jun;48(3):2135-2144. [doi: 10.1007/s00068-021-01835-5] [Medline: 34997258]
20. Mackenzie SP, Carter TH, Jefferies JG, et al. Discharged but not dissatisfied: outcomes and satisfaction of patients discharged from the Edinburgh Trauma Triage Clinic. Bone Joint J 2018 Jul;100-B(7):959-965. [doi: 10.1302/0301-620X.100B7.BJJ-2017-1388.R2]
21. Hoek AE, Anker SCP, van Beeck EF, Burdorf A, Rood PPM, Haagsma JA. Patient discharge instructions in the emergency department and their effects on comprehension and recall of discharge instructions: a systematic review and meta-analysis. Ann Emerg Med 2020 Mar;75(3):435-444. [doi: 10.1016/j.annemergmed.2019.06.008] [Medline: 31439363]
22. McKeown R, Kearney RS, Liew ZH, Ellard DR. Patient experiences of an ankle fracture and the most important factors in their recovery: a qualitative interview study. BMJ Open 2020 Feb 4;10(2):e033539. [doi: 10.1136/bmjopen-2019-033539] [Medline: 32024789]
23. Davey MS, Coveney E, Rowan F, Cassidy JT, Cleary MS. Virtual fracture clinics in orthopaedic surgery - a systematic review of current evidence. Injury 2020 Dec;51(12):2757-2762. [doi: 10.1016/j.injury.2020.11.001] [Medline: 33162011]
24. Dittrich F, Back DA, Harren AK, et al. A possible mobile health solution in orthopedics and trauma surgery: development protocol and user evaluation of the Ankle Joint app. JMIR Mhealth Uhealth 2020 Feb 26;8(2):e16403. [doi: 10.2196/16403] [Medline: 32130171]
25. Houwen T, Vugts MAP, Lansink KWW, et al. Developing mHealth to the context and valuation of injured patients and professionals in hospital trauma care: qualitative and quantitative formative evaluations. JMIR Hum Factors 2022 Jun 20;9(2):e35342. [doi: 10.2196/35342] [Medline: 35723928]

26. Van Reijen M, Vriend II, Zuidema V, van Mechelen W, Verhagen EA. The implementation effectiveness of the "Strengthen your ankle" smartphone application for the prevention of ankle sprains: design of a randomized controlled trial. BMC Musculoskelet Disord 2014 Jan 7;15:2. [doi: 10.1186/1471-2474-15-2] [Medline: 24393146]
27. Timmers T, Janssen L, van der Weegen W, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323. [doi: 10.2196/15323] [Medline: 31638594]
28. Timmers T, Janssen L, Kool RB, Kremer JAM. Educating patients by providing timely information using smartphone and tablet apps: systematic review. J Med Internet Res 2020 Apr 13;22(4):e17342. [doi: 10.2196/17342] [Medline: 32281936]
29. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144. [doi: 10.2196/34144] [Medline: 35311682]
30. Lau AYS, Piper K, Bokor D, Martin P, Lau VSL, Coiera E. Challenges during implementation of a patient-facing mobile app for surgical rehabilitation: feasibility study. JMIR Hum Factors 2017 Dec 7;4(4):e31. [doi: 10.2196/humanfactors.8096] [Medline: 29217504]
31. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136. [doi: 10.2196/jmir.6731] [Medline: 28450271]
32. Lupton D. Digitized health promotion: risk and personal responsibility for health and illness in the Web 2.0 era. In: To Fix Or To Heal: Patient Care, Public Health, and the Limits of Biomedicine: NYU Press; 2016:152-176. [doi: 10.18574/nyu/9781479878246.003.0007]
33. Risling T, Martinez J, Young J, Thorp-Froslie N. Evaluating patient empowerment in association with eHealth technology: scoping review. J Med Internet Res 2017 Sep 29;19(9):e329. [doi: 10.2196/jmir.7809] [Medline: 28963090]
34. Geerdink TH, Geerdink NJ, van Dongen JM, et al. Cost-effectiveness of direct discharge from the emergency department of patients with simple stable injuries in the Netherlands. Trauma Surg Acute Care Open 2021;6(1):e000763. [doi: 10.1136/tsaco-2021-000763] [Medline: 34722930]

Abbreviations

DD: direct discharge
ED: emergency department
GRAMMS: Good Reporting on a Mixed Methods Study
REDCap: Research Electronic Data Capture
VFC: Virtual Fracture Care

Edited by A Kushniruk; submitted 25.09.23; peer-reviewed by C Tang, EVDV Isselt; revised version received 28.06.24; accepted 08.11.24; published 14.03.25.

Please cite as:
Spierings J, Willinge G, Kokke M, Repping S, de Lange W, Geerdink T, van Veen R, van der Velde D, Goslings C, Twigt B, Collaboration Group
Patient Experiences With a Mobile Self-Care Solution for Low-Complex Orthopedic Injuries: Mixed Methods Study
JMIR Hum Factors 2025;12:e53074
URL: https://humanfactors.jmir.org/2025/1/e53074
doi:10.2196/53074

Effect of SMS Ward Round Notifications on Inpatient Experience in Acute Medical Settings: Retrospective Cohort Study

Jongchan Lee[1], MD; Soyeon Ahn[2], PhD; Jung Hun Ohn[1], MD, PhD; Eun Sun Kim[1], MD, PhD; Yejee Lim[1], MD, PhD; Hye Won Kim[1], MD, PhD; Hee-Sun Park[1], MD; Jae Ho Cho[3], MD; Sun-wook Kim[1], MD; Jiwon Ryu[1], MD; Jihye Kim[1], MD; Hak Chul Jang[1], MD, PhD; Nak-Hyun Kim[1], MD, PhD

[1]Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, 82, Gumi-ro 173 Beon-gil, Bundang-gu, Gyeonggi-do, Seongnam-si, Republic of Korea

[2]Department of Statistics, Seoul National University Bundang Hospital, Seoul National University College of Medicine, Seongnam-si, Republic of Korea

[3]Department of Hospital Medicine, Seoul St. Mary's Hospital, Seoul, Republic of Korea

Corresponding Author:
Nak-Hyun Kim, MD, PhD
Department of Internal Medicine, Seoul National University Bundang Hospital, Seoul National University College of Medicine, 82, Gumi-ro 173 Beon-gil, Bundang-gu, Gyeonggi-do, Seongnam-si, Republic of Korea

Abstract

Background: Ward rounds are an essential component of inpatient care. Patient participation in rounds is increasingly encouraged, despite the occasional complicated circumstances, especially in acute care settings.

Objective: This study aimed to evaluate the effect of real-time ward round notifications using SMS text messaging on the satisfaction of inpatients in an acute medical ward.

Methods: Since January 2021, a service implementing real-time ward round notifications via text messaging (WR-SMS) has been operational at a tertiary-care medical center in Korea. To assess its impact, we conducted a retrospective cohort study of patients admitted to the acute medical unit who participated in a patient experience survey. Patient satisfaction was compared between patients admitted in 2020 (pre–WR-SMS group) and 2021 (post–WR-SMS group).

Results: From January 2020 to December 2021, a total of 100 patients were enrolled (53 patients in the pre–WR-SMS group and 47 patients in the post–WR-SMS group). Compared with the pre–WR-SMS group, the post–WR-SMS group showed significantly greater satisfaction about being informed about round schedules (mean 3.43, SD 0.910 vs mean 3.89, SD 0.375; $P<.001$) and felt more emotionally supported during admission (mean 3.49, SD 0.800 vs mean 3.87, SD 0.397; $P<.001$). Regarding other questionnaire scores, the post–WR-SMS group showed an overall, although statistically insignificant, improvement compared with the pre–WR-SMS group.

Conclusions: Real-time round notifications using a user-friendly SMS may improve inpatient satisfaction effectively.

(JMIR Hum Factors 2025;12:e57470) doi:10.2196/57470

KEYWORDS

rounds; round-time notification; text messaging; patient experience assessment; patient experiences; patient-centeredness; patient participation

Introduction

Inpatient ward rounds constitute a salient core activity of the daily care of hospitalized patients and present a critical opportunity to deliver patient-centered care [1,2]. Ward rounds are a complex process that involves organizing the clinical care of inpatients, including diagnostic assessment and therapeutic planning by the care team [3]. For the patient, inpatient ward rounds provide an opportunity for direct patient-clinician communication, sharing of information, and participation in their own care planning [3,4]. Despite the widely recognized importance of ward rounds, an environment conducive to conducting such rounds and effectively communicating with patients is not always achievable, resulting in ineffective communication with and dissatisfaction of patients [5]. Especially in acute care settings, the participation of patients in ward rounds is likely to be limited not only due to their acute illness and fatigue [6] but also by the acute care context itself and time constraints of the ward staff [7,8].

With the emphasis on patient-centeredness, the Korean Ministry of Health and Welfare (MOHW) and the Health Insurance Review and Assessment Service (HIRA) developed the Korean Patient Experience Survey, to gain insight into the needs of inpatients and improve inpatient care quality. This survey has

been conducted biennially by telephone since 2017 among patients discharged from hospitals, and the results have repeatedly revealed that patients remain least satisfied with inpatient ward rounds [9,10]. In line with global trends, studies from the United States and the United Kingdom highlight the importance of improving patient communication and providing clear, predictable information to enhance patient satisfaction [11-13]. Furthermore, evidence suggests that interventions focusing on real-time communication [14], such as the use of SMS text messaging, can help alleviate patient dissatisfaction by improving the timeliness and transparency of ward rounds.

As part of the efforts to improve patients' satisfaction, Seoul National University Bundang Hospital (SNUBH), a 1300-bed tertiary referral hospital in the Republic of Korea, has implemented the real-time ward round notifications via text messaging (WR-SMS) service since January 2021 to inform the patients and their carers about the start of daily ward rounds, and thereby alleviating arbitrary waiting for rounds.

In 2015, a hospitalist-run acute medical unit (AMU) was established to enhance patient safety and care efficiency for patients with acute medical conditions. Although the AMU improved patient outcomes and efficiency of care [15,16], patients' perceptions of inpatient care were found to be inconsistent, which could be attributed to the acuteness and complexity of clinical conditions, the hectic hospital environment, and heightened levels of anxiety.

As mobile devices continue to advance, SMS text messaging has been suggested to enhance clear and efficient communication in various clinical contexts [14,17]. However, there remains a gap in research regarding the potential impact of informing patients about round times via SMS text messages on communication effectiveness and patient satisfaction. This study aimed to determine the effect of real-time WR-SMS service on the satisfaction of AMU inpatients.

Methods

Study Design and Clinical Setting

This retrospective cohort study was conducted at the AMU of SNUBH, a hospitalist-run 46-bed ward caring for patients with acute medical conditions, who were admitted from the emergency or outpatient departments for active acute medical care, unless being hemodynamically unstable requiring invasive monitoring and critical care or terminally ill with cancer requiring only palliative care. After receiving active acute care, patients were either discharged or transferred to a specialty ward for further treatment. The AMU was operated 24 hours a day and 7 days a week by 10 board-certified medical hospitalists on a weekly rotation basis.

The SNUBH patient-experience survey has been routinely performed by telephone since 2006. Patients eligible for the survey were randomly selected within 2 to 56 days (8 wk) after discharge. Patients unable to communicate effectively due to conditions resulting in severe cognitive or communication impairments were excluded from the survey. Following the introduction of the official Korean Patient Experience Survey developed by the Korean MOHW/HIRA in 2017, the SNUBH patient-experience survey used the very same questionnaire to assess patients' hospital experiences.

Study Participants

Adult patients ≥19 years of age who were admitted to the AMU of SNUBH between January 2020 and December 2021 for ≥1 day and who participated in the SNUBH patient-experience survey after discharge were included. Their demographic data were retrospectively collected from the electronic medical record system, and their satisfaction was assessed based on the SNUBH patient-experience survey results. To determine whether patient satisfaction improved, we compared the scores from 2020 (before the implementation of the WR-SMS, pre–WR-SMS group) with those of 2021 (after the implementation of the WR-SMS post–WR-SMS group).

Real-Time Ward Round Notification Service With Text Messaging

The WR-SMS service was first developed in September 2020. It was actively implemented in January 2021, after an introductory period of 3 months (October to December 2020) during which the staff was educated about and encouraged to use the service via campaigns and email or SMS notifications.

The WR-SMS service was embedded into the SNUBH electronic health record system, enabling the doctors to send short text messages (80 to 160 bytes; ie, SMS) to their inpatients directly from the electronic health record system with a few clicks. From their patient list, doctors were able to select or deselect the patients who would receive the SMS text message. The default content of the SMS text message read, "Ward rounds by Dr. OOO will begin shortly. Visiting times may vary depending on the room." Further adjustments to the content of the message could be made as needed by the doctor. The notification could also be sent simultaneously to other relevant medical staff. The doctors were encouraged to send this WR-SMS to their patients just before starting their ward rounds, usually right after reviewing the patients' records.

Patient-Experience Survey

The SNUBH patient-experience survey questionnaire, which uses the same questions as the official Korean Patient Experience Survey developed by the Korean MOHW/HIRA, comprises 19 questions organized into 5 domains about the patients' hospital experience (4 questions pertaining to "Services from nurses," 4 pertaining to "Services from physicians," 5 pertaining to "Medication and treatment processes," 2 pertaining to "Hospital environment," 4 pertaining to "Ensuring patients' rights", and 4 assessing patients' demographic characteristics) (detailed questionnaire is presented in Multimedia Appendix 1). Each question was scored on a 4-point Likert scale (1=very dissatisfied, 2=dissatisfied, 3=satisfied, and 4=very satisfied) to assess patients' subjective satisfaction with the quality of inpatient services. Except for the question about providing information about plans after discharge, answers were either "Yes" or "No." For this study, the responses to the "Services from physicians" and "Medication and treatment processes" domain, which are closely related to inpatient ward rounds, were considered.

Statistical Analysis

Data reflecting patients' baseline characteristics were expressed as a median with ranges for continuous variables and as frequencies and percentages for categorical variables. Data from Likert scales were expressed as means and SDs, as well as median with ranges (IQRs). Comparisons were performed using the Mann-Whitney *U* test, Student *t* test, and χ^2 test. All tests were 2-sided and performed at a significance level of .05. Statistical analyses were performed using IBM SPSS v. 21.0 (IBM Corp).

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki, revised in 2013. The SNUBH institutional review board approved this study design and waived the need for obtaining informed consent from the participants due to the retrospective design (B-2203-746-101). Participant confidentiality and anonymity were strictly maintained throughout the research process, and data were handled in a secure manner to protect their privacy. In addition, the study ensured that no undue harm or burden was placed on participants and that their rights, safety, and well-being were prioritized at all times.

Results

Baseline Characteristics of the Participants

A total of 100 patients were enrolled from January 2020 to December 2021 (53 in the pre–WR-SMS group before WR-SMS implementation and 47 in the post–WR-SMS group after WR-SMS implementation). The median age of the patients was 59.5 (range 25 - 81) years, and of the total, 59 were female. A total of 74 patients were hospitalized because of malignancy and 26 patients were admitted for non-cancer illnesses. A total of 44 and 56 patients were admitted from the emergency and outpatient departments, respectively. The median length of hospital stay was 7 (range 3 - 30) days. No statistically significant difference of patients' baseline characteristics between the pre- and post–WR-SMS groups was observed (Table 1).

Table . Baseline demographic and clinical characteristics of study participants.

Characteristics		Pre–WR-SMS group[a] (n=53)	Post–WR-SMS group[b] (n=47)	*P* value
Age (year), median (range)		58 (25-81)	60 (38-76)	.90
Sex, n (%)				.36
	Male	19 (36)	22 (47)	
	Female	34 (64)	25 (53)	
Disease, n (%)				.43
	Malignancy	37 (70)	37 (79)	
	Non-cancer illness	16 (30)	10 (21)	
Hospitalization route, n (%)				.94
	Emergency department	24 (45)	20 (43)	
	Outpatient department	29 (55)	27 (57)	
Education, n (%)				.25
	≤High school	29 (55)	32 (68)	
	≥College	24 (45)	15 (32)	
Length of stay (days), median (range)		7 (3-30)	7 (3-23)	.76

[a]Pre–WR-SMS group: patients hospitalized at the acute medical unit in 2020, before the implementation of ward round notifications via text messaging service.

[b]Post–WR-SMS group: patients hospitalized at the acute medical unit in 2021, after the implementation of ward round notifications via text messaging.

Effects of Real-Time Round Notification on Satisfaction of Hospitalized Patients

After the WR-SMS implementation in 2021, the score for the question pertaining to whether the patient had received information about round schedules improved significantly compared with 2020. Differences in the Likert scale for this question are illustrated in Figure 1. The proportion of patients who assigned a rating of 4 points (very satisfied) increased from 66% (35/53) in the pre–WR-SMS group to 92% (43/47) in the post–WR-SMS group, while the proportion of those who assigned a rating of 1 point (very dissatisfied) decreased from 6% (3/53) in the pre–WR-SMS group to 0% in in the post–WR-SMS group.

Figure 1. Changes in Likert scale scores regarding the provision of patient information on round schedules before and after the implementation of ward round notifications via text messaging. Pre–WR-SMS group: patients hospitalized at the acute medical unit in 2020, before the implementation of ward round notifications via text messaging service; Post–WR-SMS group: patients hospitalized at the acute medical unit in 2021, after the implementation of ward round notifications via text messaging service.

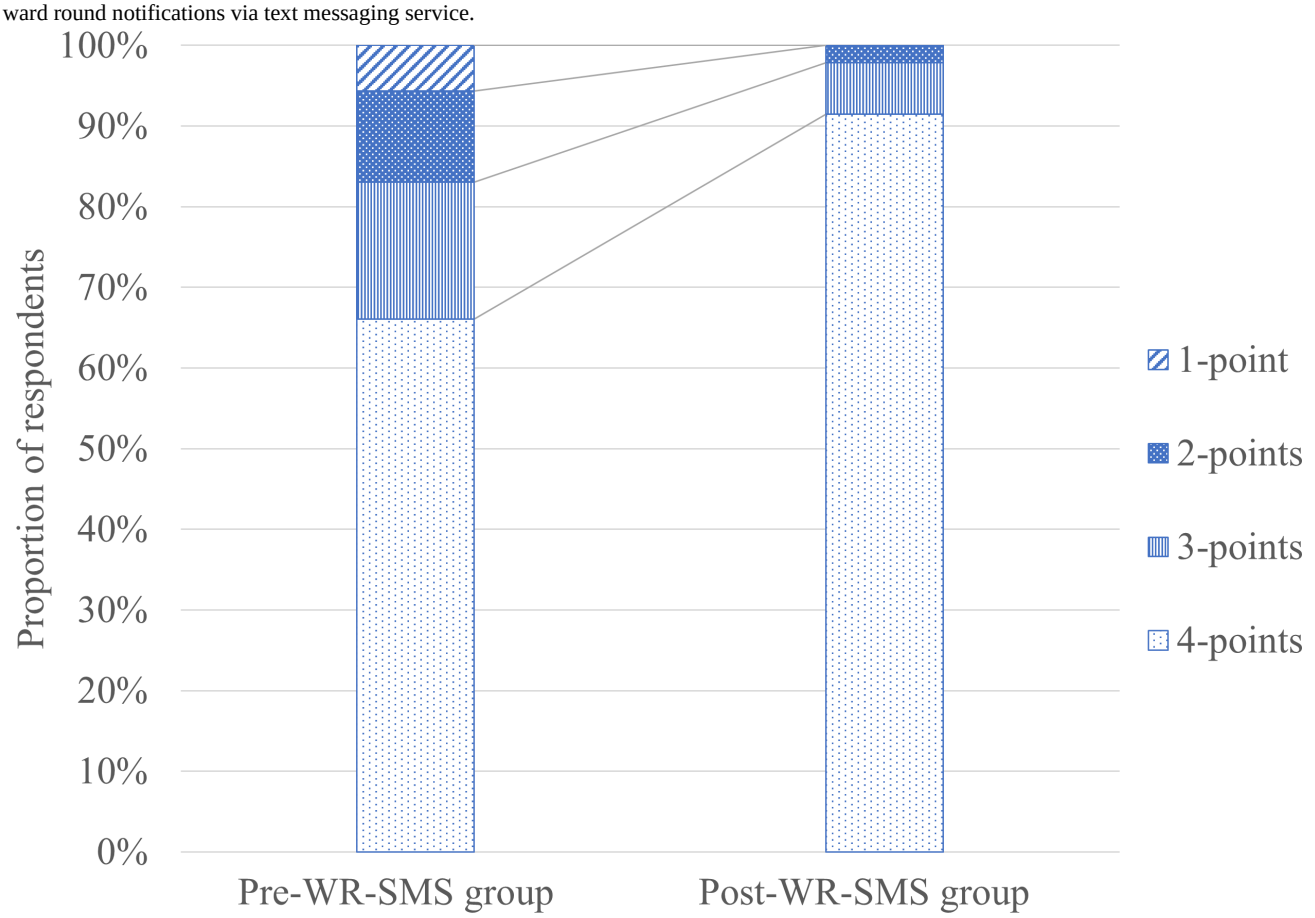


Table 2 summarizes the responses for the "Services from physicians" and "Medication and treatment processes" domains. Naturally, patients in the post–WR-SMS group showed significantly higher satisfaction about being informed about round schedules (mean 3.43, SD 0.910 vs mean 3.89, SD 0.375; *P*<.001) compared with patients in the pre–WR-SMS group. Interestingly, patients in the post–WR-SMS group also demonstrated significantly higher satisfaction with the emotional support (comfort and empathy) they received during admission (mean 3.49, SD 0.800 vs mean 3.87, SD 0.397; *P*<.001). Although not statistically significant, they were also satisfied with the description of the treatment processes (mean 3.62, SD 0.713 vs mean 3.83, SD 0.481; *P*=.09) and the explanations of possible side effects during treatment (mean 3.60, SD 0.743 vs mean 3.85, SD 0.465; *P*=.05). For all other questions in the "Services from physicians" and "Medication and treatment processes" domain, the post–WR-SMS group recorded a higher overall score compared to the pre–WR-SMS group, although it was not statistically significant (Multimedia Appendix 2).

Table . Survey results on patient satisfaction before and after the implementation of ward round notification using SMS text messaging service[a].

Parameter		Pre–WR-SMS group[b] (n=53)	Post–WR-SMS group[c] (n=47)	*t* test (df)	*P* value
Services from physicians					
	Courtesy and respect for patients			–0.93 (83.47)	
	Median (IQR)	4 (4-4)	4 (4-4)		.65
	Mean (SD)	3.81 (0.557)	3.89 (0.312)		.36
	Careful listening to patients			–0.47 (88.02)	
	Median (IQR)	4 (4-4)	4 (4-4)		.88
	Mean (SD)	3.83 (0.545)	3.87 (0.337)		.64
	Opportunities for patients to meet their doctors			–1.58 (94.08)	
	Median (IQR)	4 (3-4)	4 (4-4)		.15
	Mean (SD)	3.58 (0.663)	3.77 (0.476)		.12
	Receipt of information about round schedules			–3.37 (70.90)	
	Median (IQR)	4 (3-4)	4 (4-4)		.002
	Mean (SD)	3.43 (0.910)	3.89 (0.375)		<.001
Medication and treatment process					
	Detailed description of treatment processes			–1.72 (91.73)	
	Median (IQR)	4 (3-4)	4 (4-4)		.07
	Mean (SD)	3.62 (0.713)	3.83 (0.481)		.09
	Easy-to-understand explanation of potential side effects			–2.02 (88.60)	
	Median (IQR)	4 (3-4)	4 (4-4)		.045
	Mean (SD)	3.60 (0.743)	3.85 (0.465)		.047
	Appropriate pain relief management			–1.15 (61.49)	
	Median (IQR)	4 (4-4)	4 (4-4)		.23
	Mean (SD)	3.86 (0.462)	3.95 (0.216)		.26
	Comfort and sympathy regarding illness			–3.08 (78.06)	
	Median (IQR)	4 (3-4)	4 (4-4)		.003
	Mean (SD)	3.49 (0.800)	3.87 (0.397)		<.001
	Receipt of information on postdischarge care, n (%)	48 (90.6)	42 (89.4)	—[d]	.89

[a]Values in the first row for each item represent medians with IQRs and *P* values from the Mann-Whitney test. Values in the second row for each item represent means with SDs and *P* values from the Student *t* test (2-tailed).

[b]Pre–WR-SMS group: patients hospitalized at the acute medical unit in 2020, before the implementation of ward round notifications via text messaging service.

[c]Post–WR-SMS group: patients hospitalized at the acute medical unit in 2021, after the implementation of ward round notifications via text messaging service.

[d]Not applicable.

Discussion

Principal Findings

The WR-SMS implementation resulted in significant improvement of inpatient satisfaction among patients hospitalized in the AMU. Besides the direct improvement in scores pertaining to the receipt of round schedule information (mean 3.43, SD 0.910 vs mean 3.89, SD 0.375; *P*<.001), patients felt more emotionally supported and involved in the care process (mean 3.49, SD 0.800 vs mean 3.87, SD 0.397; *P*<.001), and also expressed higher satisfaction with the explanations provided

about treatment side effects (mean 3.60, SD 0.743 vs mean 3.85, SD 0.465; *P*=.047). These significant findings emphasize the value of real-time communication on both procedural and emotional aspects of care. Our results indicate that by enabling patients to participate more actively in routine ward rounds, WR-SMS fosters a greater sense of involvement and satisfaction, reaffirming the critical role of patient-centered communication in inpatient care.

Patient-centeredness is an essential component of high-quality health care, contributing to improved experience and clinical outcomes [18]. It is defined by the Institute of Medicine as the establishment of a partnership between the patient and health care providers to ensure that the wishes, needs, and preferences of the former are respected in the shared decision-making process [19]. Since the early 2000s, many countries have implemented patient-reported experience measures (such as standardized surveys) at the national level. The data are publicly reported, in order to assess and monitor patients' health care experiences and improve the quality of care [11,12]. For example, the United States Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS) survey, first implemented in 2006, measures patients' perspectives on the quality of care they receive during their hospital stay, with its scores influencing hospital payment [13]. The United Kingdom and the Netherlands also conduct national-level surveys to assess patient satisfaction, which are used for various purposes related to improving health care services [20,21]. Previous studies revealed that quality improvement activities were initiated after patient care performance data were publicly released [22], with even hospitals that exhibited high levels of patient satisfaction delivering higher-quality clinical care [13].

In South Korea, the Korean Patient Experience Survey has been regularly conducted by HIRA biennially since 2017 by telephone among patients discharged from hospitals. It was developed based on the American HCAHPS, but further modified to suit the Korean context and to promote improvement of current Korean health care service quality [23]. A 2015 survey by the Korea Institute for Health and Social Affairs revealed that 30.5% of hospitalized patients claimed not to have been adequately informed about ward rounds and 20.4% reported not being included in the care process [24]. Furthermore, qualitative research during the Korean Patient Experience Survey development process revealed that inpatients were frequently unsatisfied with the communication with their doctors, especially regarding the lack of advance notice and unpredictability of ward round times [23]. To gauge these concerns, a direct question about whether the patient has been informed about the ward round time was included in the official Korean Patient Experience Survey.

Despite the increasing emphasis on patient-centeredness, patients are often still excluded from their own care process [5,6]. The results of the previous Korean Patient Experience Surveys (first to third, conducted from 2017 to 2021) reveal that the scores for the "Ensuring patients' rights" and "Services from physicians" domains remain lowest [9,10,25], suggesting that communication between patients and doctors remains unsatisfactory and patients do not feel sufficiently involved in the care process. Among several factors influencing patient participation and communication, such as the ward staff's attitude, time constraints [7], and health literacy [26], the care environment context itself has been identified as critical [8]. Previous Korean patient surveys revealed that patients felt dissatisfied and even frustrated by the medical staff's impolite, indifferent, or authoritarian attitudes and faced difficulties in communicating with their doctors who always appeared busy, and took issue with unpredictable and irregular round times [23]. Particularly in acute care environments, the patients' ability to participate is further diminished by their vulnerability, illness, and anxiety; furthermore, even the ward atmosphere is often disorderly and far from suitable for effective communication. In such situations, patient involvement is notably infrequent [6].

Patients in our AMU were mostly afflicted with complex, acute illness and frequently in need of interventional procedures, imaging studies, and even surgical procedures. Therefore, these patients and their carers often missed routine ward rounds during planned procedures or did not notice that ward rounds were taking place. Furthermore, the acuteness of the patients' conditions and lack of awareness about ward rounds likely led to frustration, anxiety, and dissatisfaction of patients and their carers toward their care. Real-time notifications provided through the WR-SMS addressed these issues by increasing awareness and predictability of ward rounds, which may help mitigate the emotional burden associated with the uncertainty of acute care. This aligns with studies suggesting that structured and predictable communication mechanisms, such as SMS notifications, create a more conducive environment for patient participation in care processes. A study conducted by Redley et al [6] at an acute inpatient ward indicated that only 18% (20/52) of rounds involved patient participation in clinical decisions, although even the patients who participated in rounds indicated low preferences for participation upon arrival at the hospital. The authors suggest that the patients' preferences for participation could change during the course of their hospital stay and it is also likely that passive participation may have been driven by limited opportunities to participate [6]. Similarly, Walton et al [27] revealed that familiarity with the hospital environment tended to positively affect participation in ward rounds, with several participants highlighting the uncertainty regarding the timing of rounds as a source of difficulty and anxiety.

With the development of communication devices such as cellular phones and smartphones, text messaging via SMS is increasingly used to enhance effective communication with patients and carers [14,17]. In our institution, the WR-SMS was developed and introduced to inform inpatients about the start of ward rounds in real time to improve their opportunities to participate in them. As physicians typically review their patients' medical records right before they begin ward rounds, the WR-SMS was embedded into our electronic health record system to enhance its use instead of using a separate platform. Following its implementation in January 2021, its utilization rate rapidly increased with long-term promotional activities, reaching an average of 65% in 2021. With its use, although only a statistically insignificant increase in patients' perceptions of opportunities to meet their doctors was noted (possibly owing

to the small sample size and the nature of the ward, which is staffed by hospitalists who are primarily based within the unit), inpatient satisfaction significantly improved with regard to "providing information about round-time" and the "Medication and treatment processes" domain. However, scores for other aspects in the "Services from physicians" domain did not improve significantly, suggesting that while real-time SMS notifications enhance patients' involvement and satisfaction with specific aspects of care, they may need to be supplemented with other interventions to address broader communication gaps.

Limitations

To our knowledge, this is the first study to investigate the effect of real-time round notification via SMS on inpatients' satisfaction. Nevertheless, it has some limitations. First, it was conducted at a single referral center in South Korea. As of 2019, the penetration rate of mobile phones in South Korea was 100%, with smartphones accounting for 95%. Therefore, our results may not be reproduced in other countries with lower mobile phone penetration rates. Second, as we used information from subjects who participated in the SNUBH patient-experience survey, instead of conducting a new survey, the number of patients included in this study is relatively small, precluding further adjustments and analyses based on clinical characteristics or comorbidities. However, the clinical setting remained constant throughout the study period and the subjects for the SNUBH patient-experience survey were randomly selected, which should minimize bias. Third, since this study included only patients capable of communication as survey respondents, it was not possible to assess the impact of WR-SMS on caregiver satisfaction.

Conclusion

Effective communication between physicians and patients, particularly during inpatient ward rounds, is a crucial aspect of inpatient care. This study proposes that real-time round notifications delivered through a user-friendly SMS service could improve inpatient satisfaction, likely by providing predictability and fostering an improved perception of emotional support. In the contemporary landscape, with personal communication devices becoming increasingly ubiquitous, the strategic utilization of such devices is anticipated to significantly influence patient satisfaction.

Acknowledgments

We would like to thank Editage for the English-language editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire that was used for the Seoul National University Bundang Hospital patient-experience survey and the official Korean patient-experience survey developed by the Korean Ministry of Health and Welfare/Health Insurance Review and Assessment Service.

[DOCX File, 18 KB - humanfactors_v12i1e57470_app1.docx]

Multimedia Appendix 2

Changes in Likert scale survey scores on patient satisfaction before and after the implementation of ward round notifications via text messaging. Pre–WR-SMS group: patients hospitalized at the acute medical unit in 2020, before the implementation of ward round notifications via text messaging service; Post–WR-SMS group: patients hospitalized at the acute medical unit in 2021, after the implementation of ward round notifications via text messaging service.

[PNG File, 758 KB - humanfactors_v12i1e57470_app2.png]

References

1. Hulland O, Farnan J, Rabinowitz R, et al. What's the purpose of rounds? A qualitative study examining the perceptions of faculty and students. J Hosp Med 2017 Nov;12(11):892-897. [doi: 10.12788/jhm.2835] [Medline: 29091976]
2. Stickrath C, Noble M, Prochazka A, et al. Attending rounds in the current era: what is and is not happening. JAMA Intern Med 2013 Jun 24;173(12):1084-1089. [doi: 10.1001/jamainternmed.2013.6041] [Medline: 23649040]
3. Ward Rounds in Medicine: Principles for Best Practice: Royal College of Physicians; 2012.
4. Reddin G, Davis NF, Donald KM. Ward stories: lessons learned from patient perception of the ward round. Ir J Med Sci 2019 Nov;188(4):1119-1128. [doi: 10.1007/s11845-019-01975-z] [Medline: 30712243]
5. Aronsson L, Frithiof A, Röstedal A, Rudberg C, Ekström W. The ward round: patients' perceptions of a patient-centered approach and their suggestions for improved participation. Patient Prefer Adherence 2022;16:2203-2211. [doi: 10.2147/PPA.S343955] [Medline: 36003801]
6. Redley B, McTier L, Botti M, et al. Patient participation in inpatient ward rounds on acute inpatient medical wards: a descriptive study. BMJ Qual Saf 2019 Jan;28(1):15-23. [doi: 10.1136/bmjqs-2017-007292] [Medline: 29475980]

7. Moore L, Britten N, Lydahl D, Naldemirci Ö, Elam M, Wolf A. Barriers and facilitators to the implementation of person-centred care in different healthcare contexts. Scand J Caring Sci 2017 Dec;31(4):662-673. [doi: 10.1111/scs.12376] [Medline: 27859459]
8. McCormack B, Karlsson B, Dewing J, Lerdal A. Exploring person-centredness: a qualitative meta-synthesis of four studies. Scand J Caring Sci 2010 Sep;24(3):620-634. [doi: 10.1111/j.1471-6712.2010.00814.x] [Medline: 21050249]
9. Medical service patient experience survey results, 2017 (1st survey). : Health Insurance Review and Assessment Service; 2018.
10. Medical service patient experience survey results, 2019 (2nd survey). : Health Insurance Review and Assessment Service; 2020.
11. Goldstein E, Farquhar M, Crofton C, Darby C, Garfinkel S. Measuring hospital care from the patients' perspective: an overview of the CAHPS Hospital Survey development process. Health Serv Res 2005 Dec;40(6 Pt 2):1977-1995. [doi: 10.1111/j.1475-6773.2005.00477.x] [Medline: 16316434]
12. Fujisawa R, Klazinga NS. Measuring patient experiences (PREMS): progress made by the OECD and its member countries between 2006 and 2016. OECD. 2017. URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2017/12/measuring-patient-experiences-prems_b6354242/893a07d2-en.pdf [accessed 2025-02-26]
13. Jha AK, Orav EJ, Zheng J, Epstein AM. Patients' perception of hospital care in the United States. N Engl J Med 2008 Oct 30;359(18):1921-1931. [doi: 10.1056/NEJMsa0804116] [Medline: 18971493]
14. Wesevich A, Key-Solle M, Kandakatla A, Feeney C, Pollak KI, LeBlanc TW. Utilization of text messages to supplement rounding communication: a randomized feasibility study. J Gen Intern Med 2022 Sep;37(12):2991-2997. [doi: 10.1007/s11606-021-07285-4] [Medline: 35212877]
15. Ohn JH, Kim NH, Kim ES, et al. An acute medical unit in a Korean tertiary care hospital reduces the length of stay and waiting time in the emergency department. J Korean Med Sci 2017 Dec;32(12):1917-1920. [doi: 10.3346/jkms.2017.32.12.1917] [Medline: 29115071]
16. Kim HJ, Kim J, Ohn JH, Kim NH. Impact of hospitalist care model on patient outcomes in acute medical unit: a retrospective cohort study. BMJ Open 2023 Aug 3;13(8):e069561. [doi: 10.1136/bmjopen-2022-069561] [Medline: 37536969]
17. Souza-Silva MVR, Domingues MLDP, Chagas VS, et al. Implementation of a text messaging intervention to patients on warfarin therapy in Brazilian primary care units: a quasi-experimental study. BMC Prim Care 2022 Mar 23;23(1):54. [doi: 10.1186/s12875-022-01647-5] [Medline: 35321654]
18. Siouta E, Olsson U. Patient centeredness from a perspective of history of the present: a genealogical analysis. Glob Qual Nurs Res 2020;7:2333393620950241. [doi: 10.1177/2333393620950241] [Medline: 32944591]
19. Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century: National Academies Press (US); 2001. [doi: 10.17226/10027] [Medline: 25057539]
20. NHS patient surveys. : NHS England & the Department of Health and Social Care; 2024. URL: https://nhssurveys.org [accessed 2025-02-26]
21. Delnoij DMJ, Rademakers J, Groenewegen PP. The Dutch consumer quality index: an example of stakeholder involvement in indicator development. BMC Health Serv Res 2010 Apr 6;10:88. [doi: 10.1186/1472-6963-10-88] [Medline: 20370925]
22. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. Ann Intern Med 2008 Jan 15;148(2):111-123. [doi: 10.7326/0003-4819-148-2-200801150-00006] [Medline: 18195336]
23. Do YK, Kim JE, Lee JY, H-y L, Cho MW. Developmental research on models to evaluate patient-centeredness. : Health Insurance Review and Assessment Service; 2015.
24. Kang HJ, Lee SK, Kim SE, Hong JS, Lee KS. Korea healthcare quality report: achievements and challenges of Korea health care system. : Korea Institute for Health and Social Affairs; 2015.
25. Medical Service Patient Experience Survey results, 2021 (3rd survey). : Health Insurance Review and Assessment Service; 2022.
26. Edwards M, Davies M, Edwards A. What are the external influences on information exchange and shared decision-making in healthcare consultations: a meta-synthesis of the literature. Patient Educ Couns 2009 Apr;75(1):37-52. [doi: 10.1016/j.pec.2008.09.025] [Medline: 19036550]
27. Walton V, Hogden A, Long JC, Johnson JK, Greenfield D. Patients, health professionals, and the health system: influencers on patients' participation in ward rounds. Patient Prefer Adherence 2019;13:1415-1429. [doi: 10.2147/PPA.S211073] [Medline: 31692552]

Abbreviations

AMU: acute medical unit
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
HIRA: Health Insurance Review and Assessment Service
MOHW: Ministry of Health and Welfare
SNUBH: Seoul National University Bundang Hospital

WR-SMS: ward round notifications via text messaging

Edited by A Kushniruk; submitted 17.02.24; peer-reviewed by A Hassan, H Poulsen, JH Lee, R Bidkar; revised version received 10.01.25; accepted 14.01.25; published 12.03.25.

Please cite as:
Lee J, Ahn S, Ohn JH, Kim ES, Lim Y, Kim HW, Park HS, Cho JH, Kim SW, Ryu J, Kim J, Jang HC, Kim NH
Effect of SMS Ward Round Notifications on Inpatient Experience in Acute Medical Settings: Retrospective Cohort Study
JMIR Hum Factors 2025;12:e57470
URL: https://humanfactors.jmir.org/2025/1/e57470
doi:10.2196/57470

Device Failures and Adverse Events Associated With Rhinolaryngoscopes: Analysis of the Manufacturer and User Facility Device Experience (MAUDE) Database

Shao-Hsuan Chang[1*], PhD; Daishi Chen[2*], PhD, MD; Chi-Sheng Chen[1], MEng; Dong Zhou[1], PhD; Lung-Kun Yeh[3,4], MD, PhD

[1]Department of Biomedical Engineering, Chang Gung University, Taoyuan, Taiwan

[2]Department of Otolaryngology, Shenzhen People's Hospital, Shenzhen, China

[3]Department of Ophthalmology, Linkou Chang Gung Memorial Hospital, No. 5, Fuxing St, Guishan Dist, Taoyuan, Taiwan

[4]College of Medicine, Chang Gung University, Taoyuan, Taiwan

[*]these authors contributed equally

Corresponding Author:
Lung-Kun Yeh, MD, PhD
Department of Ophthalmology, Linkou Chang Gung Memorial Hospital, No. 5, Fuxing St, Guishan Dist, Taoyuan, Taiwan

Abstract

Background: Rhinolaryngoscopes are one of the most widely used tools by otolaryngologists and speech-language pathologists in current clinical practice. However, there is limited data on adverse events associated with or caused by the use of rhinolaryngoscopes.

Objective: In this study, we used the Manufacturer and User Facility Device Experience (MAUDE) database with the aim of providing insights that may assist otolaryngologists in better understanding the limitations of these devices and selecting appropriate procedures for their specific clinical setting.

Methods: We characterized complications associated with the postmarket use of rhinolaryngoscope devices from the US Food and Drug Administration MAUDE database from 2016 through 2023.

Results: A total of 2591 reports were identified, including 2534 device malfunctions, 56 injuries, and 1 death, from 2016 through 2023. The most common device problem with rhinolaryngoscopes was breakage (n=1058 reports, 40.8%), followed by fluid leaks (n=632 reports, 24.4%). The third most common problem was poor image quality (n=467 reports, 18%). Other device issues included contamination or device reprocessing problems (n=127 reports, 4.9%), material deformation or wear (n=125 reports, 4.8%), and device detachment (n=73 reports, 2.8%). Of the 63 reported adverse events, the most common patient-related adverse event was hemorrhage or bleeding, accounting for 18 reports, with the root causes including material deformation or wear, breakage, wrinkled rubber, or improper operation.

Conclusions: Our study offers valuable insights for endoscopists and manufacturers to recognize potential issues and adverse events associated with the use of rhinolaryngoscopes. It emphasizes the need for improving device reliability, training, and procedural protocols to enhance patient safety during diagnostic procedures.

(JMIR Hum Factors 2025;12:e67036) doi:10.2196/67036

KEYWORDS

medical device; device malfunction; rhinolaryngoscope; adverse event; MAUDE; Manufacturer and User Facility Device Experience

Introduction

Rhinolaryngoscopes are currently available in various forms, including rigid, flexible, and video versions. They are essential tools routinely used by otolaryngologists and speech-language pathologists [1,2]. Given the substantial proportion of patients who require a rhinolaryngoscopic procedure as part of a head and neck examination, rhinolaryngoscopy has become an indispensable diagnostic procedure in otolaryngological practice [3]. Modern rhinolaryngoscopes can have a distal diameter as small as 2 mm, are equipped with lighting, are flexible, and have photo and video capabilities, enabling direct visualization of the nose, throat, and airway in diverse clinical settings, including emergency, inpatient, and outpatient scenarios. The range of indications for rhinolaryngoscopy includes visualizing polyps, tumors, and sources of epistaxis in the nasal cavity and aiding in the identification of suspected tumors or adenoidal hypertrophy [4]. In the oropharynx or laryngopharynx,

rhinolaryngoscopy can be instrumental in evaluating foreign bodies and potential airway obstructions from etiologies such as neoplasm and epiglottitis, obstructive sleep apnea, dysphagia, dysphonia, tonsillar hypertrophy, glossoptosis, laryngomalacia, and vocal fold lesions [5,6]. Rhinolaryngoscopy may also assist in assessing the severity of angioedema. An alternative visualization tool is the flexible fiber-optic laryngoscope [7]. While laryngoscopes may be less expensive and simpler to use than rhinolaryngoscopes, fiber-optic rhinolaryngoscopy provides clearer visualization and better access to the larynx anatomy. In general, imaging quality, including illumination, color fidelity, resolution, and accurate length representation, plays a pivotal role in visualizing abnormalities. Therefore, the choice of instrument is often determined by clinician preference or equipment capabilities [8].

Rhinolaryngoscopes are classified as a moderate-risk medical device and require a 510(k) submission for premarket review by the US Food and Drug Administration (FDA; regulation number 21 CFR 874.4760). Although rhinolaryngoscopy is considered a generally safe procedure with rare serious adverse events, complications such as mucosal tearing, damage to anatomic structures (particularly with rigid scopes), bleeding, and laryngospasm have been reported [9]. However, there is limited data regarding device failures for both rigid and flexible rhinolaryngoscopes, incidence rates, and adverse events associated with these devices. The FDA's Manufacturer and User Facility Device Experience (MAUDE) database serves as a repository for adverse events and malfunction reports related to medical devices, and it has been used as a data source to study device-related adverse events [10,11]. In addition, capturing user experiences and integrating them into the design during medical device development has become an essential component for ensuring patient safety and device effectiveness. In this study, we used the MAUDE database to characterize postmarket complications associated with the use of rhinolaryngoscope devices from 2016 through 2023, with the aim of providing insights that may assist otolaryngologists in better understanding the limitations of these devices and selecting appropriate procedures for their specific clinical setting.

Methods

Search and Selection

MAUDE, a comprehensive postmarket surveillance database from the FDA, was chosen because of its extensive collection of medical device reports (MDRs) from both mandatory and voluntary reporters related to FDA-cleared medical devices [11]. Mandatory reporters comprise manufacturers or device user facilities, whereas voluntary reporters include physicians, patients, or other device consumers. The MDRs from MAUDE include device malfunctions and potential patient harms associated with device failures during procedures. Therefore, the event classifications used in this analysis include device malfunctions and adverse events among patients.

The MAUDE database was queried by searching for the product category "rhinolaryngoscope (Flexible or Rigid)," which included MDRs from January 2016 through December 2023. This search encompassed terms such as "rhino-laryngo videoscope," "rhino-laryngo fiberscope," "telescope," and "single use endoscope." Each MDR was then logged in a Microsoft Excel spreadsheet. Among the reports, items that may have been incorrectly classified, such as cystoscopes and arthroscopes, were manually removed. These misclassifications accounted for only a small portion of the reports and resulted from errors made by the reporters. The MDRs underwent manual review, including assessment of the date of the event, event type (device malfunction, injury, or death), and the root causes of the adverse event. Device failures were defined as reported instances of rhinolaryngoscopes not functioning as expected during a procedure.

Data Extraction

The device failure problems were classified as (1) breakage, (2) fluid leakage, (3) poor image quality, (4) contamination, (5) material deformation, (6) device detachment, (7) unintended movement, (8) mechanical problems, (9) unidentified device or use problems, (10) device overheating, (11) improper or incorrect procedures, (12) moisture damage, (13) packing problems, and (14) electrical shorting. To minimize potential bias and ensure consistency in categorization, 2 reviewers (SHC and DC) independently screened and categorized the reported events. Adverse events among patients included all reported complications resulting in injury, harm, or death during procedures associated with rhinolaryngoscopes. Following the categorization of the failure problems and adverse events, statistical analysis was conducted. The frequency of each failure type and adverse event was calculated, and the results were expressed as the number and percentage of occurrences within each category.

Additionally, to assess the context of these device failures in clinical practice, we analyzed the US Medicare population using *Current Procedural Terminology* (*CPT*) codes associated with rhinolaryngoscope diagnostic procedures (excluding surgeries). Specifically, we selected *CPT* codes 31231 (nasal endoscopy, diagnostic; unilateral or bilateral), 31575 (laryngoscopy, flexible; diagnostic), and 92511 (nasopharyngoscopy with endoscope), which may correspond to diagnostic procedures involving rhinolaryngoscopes. The data were accessed under the "Research, Statistics, Data & Systems" section. The Centers for Medicare and Medicaid Services Part B National Summary Data File was used to obtain annual procedure data based on *CPT* codes 31231, 31575, and 92511. These codes were used to compare the number of procedure claims with the number of device failure reports, allowing us to explore the relationship between the frequency of rhinolaryngoscope use (as indicated by procedure billing) and the reported failure incidents. This comparison provide a picture of the relative incidence of device failures in clinical practice compared to the frequency of procedures conducted, shedding light on potential areas for improvement in device performance or clinical training.

Ethical Considerations

This study involved querying the FDA MAUDE database and US *CPT* codes. These activities did not involve direct interaction with human subjects, nor did they involve the collection or analysis of personal health data. Therefore, such data queries typically do not require approval from an institutional review

board. Publicly available databases like the FDA MAUDE and *CPT* codes generally do not raise ethical concerns. We analyzed publicly available data and did not collect personal data or perform clinical trials, so institutional review board approval was not required.

Results

Device-Related Problems

A total of 2591 reports were identified, including 2534 device malfunctions, 56 injuries, and 1 death among patients from 2016 through 2023. The number of reports submitted significantly increased in 2021, 2022, and 2023 (n=347, n=1147, and n=918 reports, respectively), constituting greater than 90% of all reports submitted during the study period. Device malfunctions and their incidence rates are presented in Table 1. The most common device problem with rhinolaryngoscopes was breakage (n=1058 reports, 40.8%), followed by fluid leakage (n=632 reports, 24.4%). The third most common problem was poor image quality (n=467 reports, 18%). Other device issues included contamination or device reprocessing (n=127 reports, 4.9%), material deformation or wear (n=125 reports, 4.8%), device detachment (n=73 reports, 2.8%), and unintended movement (n=55 reports, 2.2%), with the remaining issues each accounting for less than 1%. Mechanical problems, device overheating, packaging issues, and electrical shorting were relatively infrequent problems, as summarized in Table 2.

Table . Event types reported to the US Food and Drug Administration Manufacturer and User Facility Device Experience (MAUDE) database for rhinolaryngoscopes per year.

Year	Deaths (n=1), n (%)	Injuries (n=56), n (%)	Malfunctions (n=2534), n (%)
2023	0 (0)	14 (25)	918 (36.2)
2022	1 (100)	7 (12.5)	1147 (45.3)
2021	0 (0)	2 (3.6)	347 (13.7)
2020	0 (0)	10 (17.8)	37 (1.5)
2019	0 (0)	14 (25)	38 (1.5)
2018	0 (0)	2 (3.6)	21 (0.8)
2017	0 (0)	5 (8.9)	13 (0.5)
2016	0 (0)	2 (3.6)	13 (0.5)

Table . Device problems reported for rhinolaryngoscopes.

Device problems	Reports (n=2591), n (%)
Breakage	1058 (40.8)
Fluid leakage	632 (24.4)
Poor image quality; optical distortion	467 (18)
Contamination; device reprocessing problem	127 (4.9)
Material deformation; wear	125 (4.8)
Detachment of device	73 (2.8)
Unintended movement	55 (2.2)
Mechanical problem	12 (0.5)
Unidentified device or use problem	11 (0.4)
Overheating of device	10 (0.4)
Improper or incorrect procedure	6 (0.2)
Moisture damage	6 (0.2)
Packaging problem	5 (0.2)
Electrical shorting; flare or flash	4 (0.2)

We conducted a comprehensive analysis of device failures, examining variations across specific manufacturers, including Olympus, Storz, Pentax, and Gyrus. A high incidence of breakage and fluid leakage reports was associated with Pentax. In comparison, Olympus devices showed a breakage rate of 4.8% (50/1028). However, Olympus devices were also associated with the highest frequency of contamination reports, followed by Pentax. These higher failure rates could be linked to the relatively larger market share of these manufacturers. Additionally, we scrutinized the frequency of nasal endoscopies performed in the US Medicare population using *CPT* codes 31231, 31575, and 92511. In 2022, a total of 1,196,520

procedures were performed. Meanwhile, the MAUDE database reported 1155 device failures. This results in an estimated failure rate of approximately 1 per 1000 procedures.

Patient-Related Adverse Events

Adverse events or complications associated with device problems were evaluated based on the MAUDE database. Adverse events in patients in reports included all instances of complications resulting in injury, harm, or death associated with the use of rhinolaryngoscopes. Adverse events in patients were categorized as (1) hemorrhage/bleeding, (2) a foreign body in the patient, (3) burns or thermal burns, (4) discomfort or pain, (5) anaphylactic shock, (6) edema, (7) laryngospasm and dyspnea, (8) unspecified tissue injury, (9) allergic reaction, and (10) death. The specified root cause of the adverse event included material and mechanical problems, device chemical contamination, improper operation, fluid leakage, overheating, and display issues, among others.

The adverse events in patients and their specified root causes are presented in Table 3. A total of 63 reported adverse events were identified, including hemorrhage/bleeding (n=18 reports, 28%), a foreign body in the patient (n=12 reports, 19%), discomfort/pain (n=10 reports, 16%), thermal burns (n=9 reports, 14%), anaphylactic shock (n=4 reports, 6%), edema (n=3 reports, 5%), laryngospasm and dyspnea (n=3 reports, 5%), unspecified tissue injury (n=2 reports, 3%), allergic reaction (n=1 report, 2%) and death (n=1 report, 2%). The most common patient-related adverse event was hemorrhage/bleeding, accounting for 18 reports, with the root causes including material deformation or wear, breakage, wrinkled rubber, or improper operation. One death was reported, with an unspecified cause. Material breakage, material wear, and improper operation were identified as the major root causes for the reported adverse events. In 2019, 1 case of edema was attributed to improper reprocessing and device contamination with chemicals. The manufacturer's investigation determined that the customer had insufficiently wiped phthalaldehyde-based disinfectants off the endoscope during reprocessing, resulting in the possibility of anaphylactic shock. Table 4 displays the numbers of adverse events from 2016 through 2023, including thermal burns, foreign bodies in patients, and discomfort, reported in 2023.

Table . Adverse events in patients from rhinolaryngoscope device problems.

Adverse event	Root cause	Events (n=63), n (%)
Hemorrhage/bleeding	Material deformation or wear; breakage; rubber wrinkling; improper operation	18 (28)
Foreign body in patient	Detachment of device or device component; breakage; material fragmentation; wear; improper operation	12 (19)
Discomfort/pain	Material wear; breakage; material invagination; mechanical problems	10 (16)
Burns/thermal burns	Scratched material; cut or torn material; burst battery; device overheating; use of device problem	9 (14)
Anaphylactic shock	Material perforation; material cracks or holes	4 (6)
Edema	Unidentified device or use problem; allergic reaction due to improper reprocessing and device chemical contamination	3 (5)
Laryngospasm and dyspnea	Display problem/poor image; unspecified issue	3 (5)
Unspecified tissue injury	Fluid/blood leakage; material wear	2 (3)
Allergic reaction	Reprocessing agent	1 (2)
Death	Unspecified issue	1 (2)

Table . Number of adverse events based on year and type.

	Events, n							
	2023	2022	2021	2020	2019	2018	2017	2016
Hemorrhage/bleeding	0	0	0	2	12	0	3	1
Foreign body in patient	4	0	0	0	2	0	5	1
Discomfort/pain	5	2	0	1	2	0	0	0
Burns/thermal burns	3	1	1	1	0	1	1	1
Anaphylactic shock	0	0	0	4	0	0	0	0
Edema	0	0	0	0	1	0	2	0
Laryngospasm and dyspnea	0	0	0	0	3	0	0	0
Unspecified tissue injury	0	2	0	0	0	0	0	0
Allergic reaction	0	0	1	0	0	0	0	0
Death	0	1	0	0	0	0	0	0

Discussion

Principal Findings

This study examined adverse events associated with the use of rhinolaryngoscopes based on the MAUDE database, analyzing 2591 common device issues and 63 patient-related adverse events from January 2016 through June 2023. Breakage (40.83%) was the most common device issue, whereas hemorrhage (28.57%) was the most common patient-related adverse event, primarily attributed to material deformation or breakage. These findings underscore the overall low occurrence of adverse events associated with the use of rhinolaryngoscopes. Previous studies have highlighted mucosal tearing and bleeding as the most common complications, with laryngospasm occurring in less than 1% of procedures [12-14]. To prevent these complications, adequate nasal decongestion and limited force are recommended. However, our results show that improper operation or use problems frequently contributed as the root cause of adverse events, including hemorrhages, foreign bodies in patients, thermal burns, pain, and edema. Operator experience with the device may also influence the occurrence of adverse events. One customer reported a broken eyepiece on a rigid scope (report number 9610773-2023-03626). Following investigation, the manufacturer attributed the issue to wear and tear coupled with excessive force. Similarly, another report detailed a damaged lens at the distal end of a scope (report number 9610773-2023-02419). Upon examination, the manufacturer concluded that this damage stemmed from user error, improper handling, and the application of excessive force. Therefore, considerations of usability and ergonomics are imperative to enhance patient care [15-17].

Despite gathering 125 reports of material wear from the MAUDE database, which accounted for one-ninth of the total breakage reports, it is noteworthy that the majority of the reported failures were not attributed to routine wear and tear, as determined by the manufacturer's investigation. This observation underscores the importance of scrutinizing reported failures to distinguish between typical wear-related issues and potentially more serious underlying concerns. It is evident that device failures may arise from various root causes beyond routine wear and tear, which manufacturers may not consistently highlight in their manuals. For instance, in the manual for the TJF-Q180V (Olympus) endoscope, there is an acknowledgment that repeated use and reprocessing of the endoscope and its accessories can lead to gradual wear and tear. This underscores the importance of thorough understanding of and adherence to manufacturer instructions, as well as recognizing potential factors contributing to device failures, beyond typical wear-related issues.

Moreover, it was observed that many cases of device failure could be linked to improper reprocessing practices. In this study, 127 device failures caused by rhinolaryngoscope contamination were reported. Although manufacturers provide recommendations for product use and reprocessing, Biadsee et al [18] reported that less than 20% of physicians adhere to the recommended decontamination process outlined by manufacturers. Another study, from Jiang et al [19], of rhinolaryngoscope device failure due to contamination using the MAUDE database from 2013 to 2019 revealed associations with laryngeal edema, rather than infection, highlighting 1 injury resulting from improper reprocessing procedures (insufficient disinfectant removal from the endoscope). Anaphylaxis resulting from the use of phthalaldehyde-based disinfectants has been reported in cases involving the cleaning of rhinolaryngoscopes and cystoscopes [20,21]. In this case (report number 9610877-2019-00238), the risk of injury from microbial

infection may be less than the potential harm caused by the disinfectant solution itself. In response to this incident, 5 members of the otolaryngology department were identified and retrained as the personnel at the hospital did not properly reprocess the endoscope. Regarding the compatibility of disinfectants, the manufacturer of the endoscope recommends that facilities adhere strictly to the instructions provided by the disinfectant manufacturer, including specific parameters such as concentration, temperature, and exposure time. Furthermore, it is crucial to ensure thorough rinsing of internal channels, external endoscope surfaces, and components with clean water to remove any residual detergent solution, thereby minimizing the risk of adverse reactions or complications associated with disinfectant use. Our results identified 1 allergic adverse event associated with a reprocessing agent, and 3 cases of edema due to improper reprocessing or device use, confirming previous literature regarding rhinolaryngoscope contamination [22,23]. Compared to other endoscope types, such as bronchoscopes and duodenoscopes, the contamination rate of rhinolaryngoscopes is less commonly associated with patient harm or death [18]. This indicates a comparatively lower risk associated with rhinolaryngoscopes in contrast to other endoscopic examinations.

This study reviewed the MAUDE database and identified 9 cases of thermal burns in patients who underwent rhinolaryngoscope examinations over 8 years. Scope overheating was reported as a root cause, likely because of prolonged procedure times with the light source on or the brightness setting of the lamp. Several studies assessed the heat effects of endoscopes in otorhinolaryngology and recommended keeping light sources at the lowest effective intensity [24-26]. MacKeith et al [27] evaluated the amount of heat produced by endoscopes and showed that larger-diameter endoscopes attain a higher temperature. Chitnavis [28] demonstrated that even momentary proximity could cause a thermal burn to a patient's skin, without generating smoke or fire. The risk of thermal injury may be associated with the light source, endoscope caliber, and angulation [25,27,29,30]. Consequently, available cameras are often equipped with a regulation system capable of automatic gain control in poor lighting situations [31].

Furthermore, it is noteworthy that the number of MDRs underwent a significant surge in 2021, 2022, and 2023, with 347, 1147, and 918 reports, respectively. This notable increase underscores the importance of continued vigilance and thorough investigation into the factors driving these trends to ensure the safety and efficacy of medical devices. The growth in frequency can be attributed to several major factors, including the increased prevalence of laryngeal diseases, heightened use of laryngoscopes in airway management, guideline recommendations, elevated demand for respiratory products, and the impact of the COVID-19 pandemic. However, there was no significant rise in the number of nasal endoscopies performed in the US Medicare population with *CPT* codes 31231, 31575, and 92511 from 2016 to 2022. Another plausible explanation for the increased MDRs could stem from new product launches and strategic activities by manufacturers, which have had a moderate impact on the laryngoscope market, influencing the growth in the frequency of MDRs.

The statistical analysis conducted using the FDA's MAUDE database carries significant implications for clinical practice. These data not only help identify trends and patterns in device failures but also provide critical safety information for clinical health care. By analyzing the MAUDE data, we can identify failure types and frequencies associated with specific devices, such as rhinolaryngoscopes, thereby understanding their potential impact on patient safety. The data reveal the sources of device failures, including variations across manufacturers and models, which can aid clinical decision-makers in selecting devices based on more informed criteria. Furthermore, by examining the time trends of device failures in the MAUDE database, we can correlate these trends with medical policies or other external factors. This trend analysis can assist health care institutions in adjusting operational processes and training programs to address the growing issue of device failures, especially during times of infectious disease outbreaks or when new devices are introduced. The data derived provide empirical evidence that can help in developing more effective clinical guidelines and training plans, ultimately contributing to improved quality of health care services [32]. These findings highlight the importance of timely adjustments to clinical operational processes and training programs to address the growing issue of device failures and provide guidance for future device improvements and clinical recommendations.

Limitations

A major limitation of this study is its reliance on reports from the FDA's MAUDE database, which may not account for comorbidities involved in reported device failures. Furthermore, the information may be incomplete or limited because MDRs can be submitted by health care professionals, patients, or manufacturers. Reporting variations between private-practice clinics and large academic medical centers may also affect data accuracy and which brands are used. Therefore, establishing causative associations for some reported adverse events is challenging. However, it is important to note that manufacturers are required to investigate reported adverse events and device failures, and their findings or follow-up reports are included in the MAUDE database. The manufacturer reports help to mitigate potential bias arising from self-reported data. Since the MAUDE database primarily focuses on device-related issues, we recognize the possibility of underreporting or inconsistent reporting practices. Nonetheless, the inclusion of both user-reported and manufacturer-investigated information adds a level of rigor to the analysis, allowing for a more nuanced understanding of the issues related to device failures. Despite these limitations, our study provides important insights for endoscopists and manufacturers to recognize potential issues and adverse events associated with the use of rhinolaryngoscopes, emphasizing the importance of patient safety.

Conclusions

Although our findings underscore the overall low occurrence of adverse events associated with the use of rhinolaryngoscopes, the results indicate that improper operation frequently contributed to adverse events, including hemorrhage, foreign bodies in patients, thermal burns, pain, and edema. Operator

experience with the device may also influence the occurrence of adverse events. Results from this study will be important for endoscopists and manufacturers to have a thorough understanding of the equipment and its limitations. Future research should assess the broader organizational impact, including otolaryngology teams, documentation practices, clinician training, and patient perspectives.

Data Availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions

Supervision: SHC and LKY
Conceptualization: SHC and DC
Methodology: SHC and CSC
Formal analysis: SHC and DZ
Writing—original draft preparation: SHC
Writing—review and editing: DC

Conflicts of Interest

None declared.

References

1. Nankivell PC, Pothier DD. Nasal and instrument preparation prior to rigid and flexible nasendoscopy: a systematic review. J Laryngol Otol 2008 Oct;122(10):1024-1028. [doi: 10.1017/S0022215108002533] [Medline: 18492307]
2. Rosen CA, Amin MR, Sulica L, et al. Advances in office-based diagnosis and treatment in laryngology. Laryngoscope 2009 Nov;119 Suppl 2:S185-S212. [doi: 10.1002/lary.20712] [Medline: 19856405]
3. Zacharias SRC, Weinrich B, Brehm SB, et al. Assessment of vibratory characteristics in children following airway reconstruction using flexible and rigid endoscopy and stroboscopy. JAMA Otolaryngol Head Neck Surg 2015 Oct;141(10):882-887. [doi: 10.1001/jamaoto.2015.1740] [Medline: 26402578]
4. Aremu SK. Flexible nasopharyngolaryngoscopy: evaluation and appraisal of its effectiveness and diagnostic yield, the Nigerian experience. J Family Med Prim Care 2019 Oct;8(10):3399-3403. [doi: 10.4103/jfmpc.jfmpc_489_19] [Medline: 31742176]
5. Bentsianov BL, Parhiscar A, Azer M, Har-El G. The role of fiberoptic nasopharyngoscopy in the management of the acute airway in angioneurotic edema. Laryngoscope 2000 Dec;110(12):2016-2019. [doi: 10.1097/00005537-200012000-00007] [Medline: 11129012]
6. Zeleník K, Walderová R, Kučová H, Jančatová D, Komínek P. Comparison of long-term voice outcomes after vocal fold augmentation using autologous fat injection by direct microlaryngoscopy versus office-based calcium hydroxylapatite injection. Eur Arch Otorhinolaryngol 2017 Aug;274(8):3147-3151. [doi: 10.1007/s00405-017-4600-1] [Medline: 28478500]
7. Milstein CF, Charbel S, Hicks DM, Abelson TI, Richter JE, Vaezi MF. Prevalence of laryngeal irritation signs associated with reflux in asymptomatic volunteers: impact of endoscopic technique (rigid vs. flexible laryngoscope). Laryngoscope 2005 Dec;115(12):2256-2261. [doi: 10.1097/01.mlg.0000184325.44968.b1] [Medline: 16369176]
8. Cohen SM, Pitman MJ, Noordzij JP, Courey M. Evaluation of dysphonic patients by general otolaryngologists. J Voice 2012 Nov;26(6):772-778. [doi: 10.1016/j.jvoice.2011.11.009] [Medline: 22285452]
9. Wrigley SR, Black AE, Sidhu VS. A fibreoptic laryngoscope for paediatric anaesthesia. A study to evaluate the use of the 2.2 mm Olympus (LF-P) intubating fibrescope. Anaesthesia 1995 Aug;50(8):709-712. [doi: 10.1111/j.1365-2044.1995.tb06100.x] [Medline: 7645703]
10. Wilson N, Abdallah M, Jaber F, et al. Analysis of reported adverse events associated with over-the-scope endoscopic suturing system: an FDA MAUDE database study. Obes Surg 2023 Apr;33(4):1253-1258. [doi: 10.1007/s11695-023-06525-2] [Medline: 36810809]
11. Gurtcheff SE. Introduction to the MAUDE database. Clin Obstet Gynecol 2008 Mar;51(1):120-123. [doi: 10.1097/GRF.0b013e318161e657] [Medline: 18303506]
12. Ricchetti A, Becker M, Dulguerov P. Internal carotid artery dissection following rigid esophagoscopy. Arch Otolaryngol Head Neck Surg 1999 Jul;125(7):805-807. [doi: 10.1001/archotol.125.7.805] [Medline: 10406322]
13. Jones JW, Perryman M, Judge P, et al. Resident education in laryngeal stroboscopy and perceptual voice evaluation: an assessment. J Voice 2020 May;34(3):442-446. [doi: 10.1016/j.jvoice.2018.11.016] [Medline: 30545492]
14. Ngan JH, Fok PJ, Lai EC, Branicki FJ, Wong J. A prospective study on fish bone ingestion. Experience of 358 patients. Ann Surg 1990 Apr;211(4):459-462. [doi: 10.1097/00000658-199004000-00012] [Medline: 2322040]
15. Gupta SP, Pidgeon A. An analytical approach to identifying potential use-related issues concerning a medical device under development. J Med Eng Technol 2016;40(3):61-71. [doi: 10.3109/03091902.2015.1132785] [Medline: 26795586]

16. van der Peijl J, Klein J, Grass C, Freudenthal A. Design for risk control: the role of usability engineering in the management of use-related risks. J Biomed Inform 2012 Aug;45(4):795-812. [doi: 10.1016/j.jbi.2012.03.006] [Medline: 22466009]
17. Gao X, Wen Q, Duan X, et al. A hazard analysis of class I recalls of infusion pumps. JMIR Hum Factors 2019 May 3;6(2):e10366. [doi: 10.2196/10366] [Medline: 31066695]
18. Biadsee A, Crosby L, Chow W, Sowerby LJ. Cost minimization analysis of nasopharyngoscope reprocessing in community practice. J Otolaryngol Head Neck Surg 2023 Feb 8;52(1):8. [doi: 10.1186/s40463-022-00610-9] [Medline: 36750881]
19. Jiang R, Kasle DA, Alzahrani F, Kohli N, Lerner MZ. A manufacturer and user facility device experience analysis of upper aerodigestive endoscopy contamination: is flexible laryngoscopy different? Laryngoscope 2021 Mar;131(3):598-605. [doi: 10.1002/lary.28826] [Medline: 32558941]
20. Sokol WN. Nine episodes of anaphylaxis following cystoscopy caused by Cidex OPA (ortho-phthalaldehyde) high-level disinfectant in 4 patients after cytoscopy. J Allergy Clin Immunol 2004 Aug;114(2):392-397. [doi: 10.1016/j.jaci.2004.04.031] [Medline: 15316522]
21. Suzukawa M, Komiya A, Koketsu R, et al. Three cases of ortho-phthalaldehyde-induced anaphylaxis after laryngoscopy: detection of specific IgE in serum. Allergol Int 2007 Sep;56(3):313-316. [doi: 10.2332/allergolint.C-06-51] [Medline: 17582212]
22. Kramer A, Kohnen W, Israel S, et al. Principles of infection prevention and reprocessing in ENT endoscopy. GMS Curr Top Otorhinolaryngol Head Neck Surg 2015;14:Doc10. [doi: 10.3205/cto000125] [Medline: 26770284]
23. Mistry R, Russell RV, Walker N, Ofo E. The single-use rhinolaryngoscope: an evaluation and cost comparison. J Laryngol Otol 2020 Sep 28;134(9):1-8. [doi: 10.1017/S0022215120001656] [Medline: 32985399]
24. Bottrill I, Perrault DF Jr, Poe D. In vitro and in vivo determination of the thermal effect of middle ear endoscopy. Laryngoscope 1996 Feb;106(2 Pt 1):213-216. [doi: 10.1097/00005537-199602000-00020] [Medline: 8583856]
25. Lewis T, Levin M, Sommer DD. Too hot to handle-quantifying temperature variations in the nasal endoscope ocular assembly and light post. Am J Rhinol Allergy 2020 Mar;34(2):262-268. [doi: 10.1177/1945892419892182] [Medline: 31779480]
26. Yavuz Y, Skogås JG, Güllüoğlu MG, Langø T, Mårvik R. Are cold light sources really cold? Surg Laparosc Endosc Percutan Tech 2006 Oct;16(5):370-376. [doi: 10.1097/01.sle.0000213711.32805.15] [Medline: 17057587]
27. MacKeith SAC, Frampton S, Pothier DD. Thermal properties of operative endoscopes used in otorhinolaryngology. J Laryngol Otol 2008 Jul;122(7):711-714. [doi: 10.1017/S0022215107000734] [Medline: 17942007]
28. Chitnavis J. Silent burn: the hidden danger and effects of bright light from fibre-optic cables in arthroscopic knee surgery. J Surg Case Rep 2020 Apr;2020(4):rjaa068. [doi: 10.1093/jscr/rjaa068] [Medline: 32280444]
29. Prasad N, Tavaluc R, Harley E. Thermal injury to common operating room materials by fiber optic light sources and endoscopes. Am J Otolaryngol 2019;40(5):631-635. [doi: 10.1016/j.amjoto.2019.05.007] [Medline: 31178283]
30. Souza E Silva TX, Nicolau ABF, Antunes ML. Thermal variation in human temporal bone using rigid endoscope. Braz J Otorhinolaryngol 2024;90(3):101381. [doi: 10.1016/j.bjorl.2023.101381] [Medline: 38364523]
31. Modlin IM, Kidd M, Lye KD. From the lumen to the laparoscope. Arch Surg 2004 Oct;139(10):1110-1126. [doi: 10.1001/archsurg.139.10.1110] [Medline: 15492154]
32. Hoffmann C, Avery K, Macefield R, et al. Usability of an automated system for real-time monitoring of shared decision-making for surgery: mixed methods evaluation. JMIR Hum Factors 2024 Apr 10;11:e46698. [doi: 10.2196/46698] [Medline: 38598276]

Abbreviations

***CPT*:** *Current Procedural Terminology*
FDA: Food and Drug Administration
MAUDE: Manufacturer and User Facility Device Experience
MDR: medical device report

Edited by A Kushniruk; submitted 30.09.24; peer-reviewed by J Walsh, LE Galvao Martins; revised version received 11.01.25; accepted 14.01.25; published 05.02.25.

Please cite as:
Chang SH, Chen D, Chen CS, Zhou D, Yeh LK
Device Failures and Adverse Events Associated With Rhinolaryngoscopes: Analysis of the Manufacturer and User Facility Device Experience (MAUDE) Database
JMIR Hum Factors 2025;12:e67036
URL: https://humanfactors.jmir.org/2025/1/e67036
doi:10.2196/67036

Swedish Version of the System Usability Scale: Translation, Adaption, and Psychometric Evaluation

Helene Åvik Persson[1], RNCS, PhD; Charlotte Castor[1], RNCS, PhD; Nilla Andersson[1], OTR, PhD; Mia Hylén[2,3], RNCS, PhD

[1]Department of Health Sciences, Lund University, Box 117, Sölvegatan 19, Lund, Sweden

[2]Department of Care Science, Faculty of Health and Society, Malmo University, Malmö, Sweden

[3]Department of Intensive and Perioperative Care, Skane University Hospital, Malmö, Sweden

Corresponding Author:
Helene Åvik Persson, RNCS, PhD
Department of Health Sciences, Lund University, Box 117, Sölvegatan 19, Lund, Sweden

Abstract

Background: The Swedish health care system is undergoing a transformation. eHealth technologies are increasingly being used. The System Usability Scale is a widely used tool, offering a standardized and reliable measure for assessing the usability of digital health solutions. However, despite the existence of several translations of the System Usability Scale into Swedish, none have undergone psychometric validation. This highlights the urgent need for a validated and standardized Swedish version of the System Usability Scale to ensure accurate and reliable usability evaluations.

Objective: The aim of the study was to translate and psychometrically evaluate a Swedish version of the System Usability Scale.

Methods: The study utilized a 2-phase design. The first phase translated the System Usability Scale into Swedish and the second phase tested the scale's psychometric properties. A total of 62 participants generated a total of 82 measurements. Descriptive statistics were used to visualize participants' characteristics. The psychometric evaluation consisted of data quality, scaling assumptions, and acceptability. Construct validity was evaluated by convergent validity, and reliability was evaluated by internal consistency.

Results: The Swedish version of the System Usability Scale demonstrated high conformity with the original version. The scale showed high internal consistency with a Cronbach α of .852 and corrected item-total correlations ranging from 0.454 to 0.731. The construct validity was supported by a significant positive correlation between the System Usability Scale and domain 5 of the eHealth Literacy Questionnaire (P=.001).

Conclusions: The Swedish version of the System Usability Scale demonstrated satisfactory psychometric properties. It can be recommended for use in a Swedish context. The positive correlation with domain 5 of the eHealth Literacy Questionnaire further supports the construct validity of the Swedish version of the System Usability Scale, affirming its suitability for evaluating digital health solutions. Additional tests of the Swedish version of the System Usability Scale, for example, in the evaluation of more complex eHealth technology, would further validate the scale.

Trial Registration: ClinicalTrials.gov NCT04150120; https://clinicaltrials.gov/study/NCT04150120

(JMIR Hum Factors 2025;12:e64210) doi:10.2196/64210

KEYWORDS

application; Swedish; System Usability Scale; usability; validation

Introduction

In the rapidly evolving landscape of global health care, the advent of eHealth technologies has emerged as a transformative force that promises innovative solutions to the multifaceted challenges faced by health care systems worldwide [1,2]. The Swedish health care system is currently transforming along these lines. The use of digital applications and other digital contact methods, collectively described under the term eHealth, is increasing. The World Health Organization (WHO) defines eHealth as "the use of information and communication technologies (ICT) for health." These technologies include a wide range of systems interventions, applications, and devices such as mobile health and telehealth [3]. There is compelling evidence for the increasing influence of eHealth on the provision of health care globally today and how it is enhancing the efficiency and responsiveness of health systems to meet people's needs and expectations [3]. It is essential to ensure that health technologies are designed appropriately to meet the needs of end users before deploying them as health interventions [4].

Employing robust evaluation methods to ensure high-level usability has been recognized as a crucial component of good practice for achieving this goal [5].

Usability is defined 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" [6]. To determine the potential usability of a digital solution, the System Usability Scale (SUS) has been widely adopted as a standardized evaluative device. In 1996, Brooke [7] published the SUS as an instrument that could easily measure usability. Since then, it has been used to evaluate computer systems, applications, and other digital solutions in a wide range of areas [8-10]. The SUS is a questionnaire, consisting of 10 items each scored on a 5-point Likert scale from "strongly disagree" to "strongly agree." It is easy to administer and has been shown to generate results with good reliability and validity [7,8,11]. The instrument is free, and no fees are required to use it [7]. With an international reach, it is the most widely used standardized questionnaire for the assessment of perceived usability today [10]. The SUS can be used to evaluate a wide range of usability in products, including digital applications, mobile devices, and web pages [8-10].

However, as psychometric properties are sample dependent, it is essential to evaluate the psychometric properties when using patient-reported outcome measure in new settings or populations [12]. The SUS has been translated into numerous languages such as Chinese [13], Finnish [14], French [15], Hindi [15], Indonesian [16], and Polish [17]. It has undergone psychometric validation [18], including in Arabic [19], Danish [20], Dutch [21], German [22], Italian [23], Malay [24], Persian [25], Portuguese [26], Slovene [27], and Spanish [28]. The psychometric properties that have resulted from these studies show that adapted versions of the SUS are a reliable tool for usability assessments. However, a number of these studies adopted a general focus and examined only the total sum in the test. Only a small number of studies have tested the instrument at an item level [19-28], with none of them in Swedish. This emphasizes the need for a comprehensive testing of a Swedish version of the SUS on an item level. Determining the robustness is critical for ensuring that the measurement instrument has sufficient validity in the proposed context.

Lewis [10] outlines the essential requirements for translating the SUS into multiple languages and conducting validations across diverse countries. In Sweden, one translation of the SUS has been published in a scientific journal [29], although several unpublished versions exist [9]. However, the linguistic discrepancies between these translations give room for ambiguities, and a united translation with a rigorous psychometric testing in a Swedish context is warranted.

In 2016, an initial vision was for Sweden to be "the best in the world" in eHealth by 2025. However, a more realistic view today is that Sweden has made significant progress in this area. Within the European context, Sweden demonstrates a distinct approach to digitalization in health care, emphasizing collaboration and innovation to address specific challenges and opportunities [30].

With this in mind, there is a great need for a nationally united SUS that has been rigorously tested and proven effective. A robust process of evaluating the psychometric properties of a Swedish SUS will foster participatory usability research and ultimately improve the quality of health care services on a broader scale. Although there are different Swedish translations of the SUS, there is still a lack of psychometric testing showing their robustness. Before any Swedish SUS can be recommended for use, both translation and psychometric evaluation of the instrument are necessary. Therefore, the aim of this study was to translate and psychometrically evaluate a Swedish version of the SUS.

Methods

Overview

This study consists of 2 steps. In step 1, the SUS was translated and adapted into Swedish. In step 2, the psychometric properties of the Swedish version were tested: data quality, scaling assumptions, acceptability, convergent validity, and internal consistency. The translation, adaptation process, and psychometric evaluation adhered to the COSMIN (Consensus-Based Standards for the Selection of Health Status Measurement Instruments) checklist [31].

The SUS Instrument

The SUS instrument consists of 10 items (statements). The items are divided equally into positively and negatively worded statements. Requests for a response are graded on a scale of agreement ranging from strongly disagree (1) to strongly agree (5). The score is calculated as follows: for positively worded items (1, 3, 5, 7, and 9), the score is the position on the scale (1-5) minus 1, and for negatively worded items (2, 4, 6, 8, and 10), the score is the position on the scale minus 5. Individual item scores, therefore, can range from 0 - 4. The sum of all 10 scores is then multiplied by 2.5, resulting in a total score ranging from 0 - 100, with higher numbers representing greater usability [7,8].

To date, factor analysis has not been able to show conclusively whether the SUS consists of one factor (usability) or two (usability and learnability) [11]. Efforts to replicate these findings have led to the conclusion that addressing the instrument as 2 dimensional has no practical or theoretical interest. This study, therefore, treats the SUS as a unidimensional instrument of perceived usability [32].

Translation and Adaption

According to Brooke's [7] original formulation of the SUS, no formal permission is needed for translation, and it can be used free of charge. Further, Brooke [7] allows for the possibility to, in any version, exchange the wording of the scale to a word or expression suitable for the situation. The translation process adopted in this study was inspired by Beaton et al [33]. First, the original instrument was translated from English to Swedish by one of the researchers (CC), proceeding from the versions presented by Bangor et al [34] and Lewis [10]. During the process, perceived difficulties and uncertainties were noted. This first version was then reviewed and discussed by the group, consisting of researchers familiar with eHealth, until consensus

was reached. Second, an authorized translator, naive to the research field, carried out a back translation on the instrument and the notes taken during the research process. Finally, the research group reviewed all versions of the translation, with all notes attached, and finalized a second version of a Swedish SUS. This version was then compared to other SUS translations within the Scandinavian countries for content validity and the original version for expressions and conceptions that could have been culturally influenced. Following this last step, the group then decided upon a final version by consensus. The response options were structured in the same way as the original questionnaire, although the phrases “strongly disagree” and “strongly agree” were exchanged for the Swedish equivalents of “totally disagree” and “totally agree.”

Psychometric Evaluation

Sample

The evaluation of the psychometric properties was carried out in conjunction with a larger intervention study evaluating eHealth, the eChildHealth tablet study [35]. Parents of children with a range of illnesses and health conditions who were patients in the pediatric department of a level-3 hospital in the south of Sweden were invited to take part. A total of 66 parents to 52 unique children gave informed consent and were included in the eChildHealth tablet study. Of these parents, 62 provided information on the SUS, resulting in a total of 82 measurements.

Data Collection

Parents were introduced to an app on a tablet computer through which they could communicate with health care staff after their child had been discharged from hospital. The app made it possible to continue to communicate with hospital staff whom the parents knew well, through chat messages, sending photos, video calls, and predesigned questionnaires. Using a questionnaire, data on various aspects of eHealth were collected for each parent. The SUS was one aspect, and the eHealth Literacy Questionnaire (eHLQ) [36] was another. A study-specific questionnaire was used to collect demographic data such as age and level of education. Between October 2022 and October 2023, the 66 parents were included in study. Data were collected after 1 - 2 weeks of use and at a second time point for those participants who used a tablet for more than 1 month. These parents constitute the eligible participants for this study. A sufficient adequate sample size of approximately 80 measurements was based on recommendations from COSMIN [31] and Beaton et al [33].

Data Analysis

The psychometric properties of the SUS were analyzed with IBM SPSS Statistics for Windows (version 28.0, IBM Corp). Descriptive statistics (mean, SD, and percentage) were used to visualize participants’ characteristics along with data quality, scaling assumptions, and acceptability. Construct validity was evaluated by convergent validity, and reliability was evaluated by internal consistency.

For data quality, use within the clinical setting was determined by item nonresponse and missing scale scores, as they reflect the acceptance and understanding of a measure [37]. Data quality was determined as high if the percentage of missing data per item was low (<10% acceptable). In this study, there were missing data in 5 items (ranging from 1/81, 1% to 3/81, 4%), representing high data quality. Participants with more than 3 unanswered questions were excluded (n=1), while mean imputation was carried out for those with 1 or 2 missing items (n=6, all missing 1 item) [38]. This resulted in a total of 81 measurements being included in the final analysis.

Regarding scaling assumptions, the dimensionality of the SUS has been evaluated previously [32]. In accordance with these studies, this study assumes that the SUS is to be treated as unidimensional (all items measure the same construct). Instruments composed of Likert-scale items can be summarized if they have similar means and SDs. Furthermore, item-total correlations (the correlation between each individual item score and the total score) would indicate if all items contribute equally to the total score. In line with Hobart et al [37], item-total correlations with values of $r \geq 0.3$ were regarded as sufficient for summing up the items to a total sum score.

To evaluate the acceptability of score distributions, ceiling and floor effects along with skewness were calculated. Ceiling and floor effects were regarded as present if they exceeded 90%, that is, the percentage of responses for the lowest and highest scores. Skewness statistics should preferably be within the range of −1 to +1 [37].

Construct validity was explored through convergent validity. It was evaluated with the correlation between the high total sum of the SUS and one of the domains of the eHLQ [36], with the hypothesis that it would correlate with the total score of the SUS. Both instruments were distributed to the parents at the same time points. The eHLQ has already been translated, adapted, and validated within a Swedish context [39]. The eHLQ consists of 7 domains across 35 items, with each domain being extractable and treatable as a separate scale. Responses to each domain on the eHLQ are recorded on a Likert scale from 1 (strongly disagree) to 4 (strongly agree). The eHLQ was designed to be used to understand and evaluate people’s interaction with digital health services [36,39]. The research group independently reviewed the eHLQ and, after discussion, decided to use domain 5, measuring the motivation to engage with digital services. The correlation between the instruments was assessed using the nonparametric correlation coefficient of Kendall Tau-b. With regard to the comparative instrument not being used in total (only 1 domain), the correlation was regarded as acceptable if it was moderate or greater (>0.3) [40] and with the level of significance being $P<.05$.

Internal consistency—how items are related to each other—was explored according to the indicators recommended by Hobart et al [37]: corrected item-total correlations and Cronbach α. The cutoffs were >.04 for acceptable corrected item-total correlations and >.8 for Cronbach α [37,41]. The internal consistency was completed with an SEM to analyze the measurement error of the instrument. The SEM represents the smallest difference in scores and indicates a change on a group level. SEM was analyzed with $SD_{baseline} \times \sqrt{1}$ – reliability and complemented with a CI 95% [12].

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki [42] and approved by the Swedish Ethical Review Authority (2021 - 05077). The invited parents were recruited through gatekeepers at the pediatric department. Information regarding the study was initially provided orally and then followed up by written information, before written informed consent was obtained by a study nurse. Data were handled confidentially, and participating parents were able to quit without any explanation or impact on the care their children received. Participation in the study was voluntary, and no financial or other form of compensation was provided to the participants.

Results

Translation and Adaption

Both translations (forward and backward) were similar and did not differ substantially. Overall, the conformity of phrasing was high, with some discrepancy for phrasing in statements 2 (I found the system unnecessarily complex) and 8 (I found the system very cumbersome to use). In statement 2, the word “complicated” was suggested in the back translation from the Swedish word “komplex.” In statement 8, the word “awkward” was suggested instead of the original word “cumbersome.”

All such discrepancies were discussed within the research group and a consensus was reached for the final version. The final step, comparing the Swedish version to other Scandinavian versions, generated no further changes. Table 1 shows the original English version and the final Swedish version of the SUS.

Table . Original English version [7] and the proposed Swedish version of the System Usability Scale, including (1) response alternatives and (2) statements.

Component	English version	Swedish version
Response alternatives	• 1 - Strongly disagree • 2 • 3 • 4 • 5 - Strongly agree	• 1 - Instämmer inte alls • 2 • 3 • 4 • 5 - Instämmer helt
Statements	1. I think that I would like to use this system frequently 2. I found the system unnecessarily complex 3. I thought the system was easy to use 4. I think that I would need the support of a technical person to be able to use this system 5. I found the various functions in this system were well integrated 6. I thought that there was too much inconsistency in this system 7. I would imagine that most people would learn to use this system very quickly 8. I found the system very cumbersome to use 9. I felt very confident using the system 10. I needed to learn a lot of things before I could get going with this system	1. Jag tror att jag skulle vilja använda denna applikation ofta 2. Jag uppfattar denna applikation som onödigt komplex 3. Jag tycker att denna applikation är enkel att använda 4. Jag tror att jag skulle behöva stöd för att kunna använda denna applikation 5. Jag upplever att de olika funktionerna i denna applikation var väl integrerade 6. Jag tycker att applikationen är inkonsekvent 7. Jag föreställer mig att de flesta personer skulle lära sig att använda denna applikation väldigt snabbt 8. Jag upplever denna applikation som krånglig 9. Jag känner mig trygg med att använda denna applikation 10. Jag behövde lära mig många saker innan jag kunde komma igång med denna applikation

[a]In this study, the Swedish instruction was as follows: "Markera det alternativ som bäst beskriver din reaktion för applikation i surfplatta idag," meaning “Please indicate your agreement with the following statements, one at a time.”

Psychometric Evaluation of the Swedish Translation

A total of 62 individuals were included in this study. Of these, 20 answered the questionnaire, including the SUS, twice. This resulted in a total of 81 measurements included in the analysis. Demographic data for the participants are shown in Table 2.

Table . Demographic data of participants included in the study (N=62).

Demographic data	Values
Age (years), median (range)	33 (22 - 52)
Gender, n (%)	
Female	41 (66)
Male	20 (32)
Unknown	1 (2)
Marital status, n (%)	
Married	33 (53)
Living together	28 (45)
Divorced or separated	1 (2)
Education level, n (%)	
High school	18 (29)
College or university	42 (68)
Other	2 (3)
Born in Sweden, n (%)	
Yes	52 (84)
No	6 (10)
Unknown	4 (6)
First language	
Swedish	56 (90)
Other	6 (10)

Data Quality

The percentage of missing data per item was low (ranging from 0/81, 0% to 3/81, 4%) across all items. A tendency could be seen for a higher percentage within the 3 highest steps of the scale, resulting in 5 items having 0% in the lowest points of the 5-point scale (Table 3).

Table . Missing data (n and %) and item frequency distribution (%) of answers per response alternative in each question of the Swedish System Usability Scale (n=81). The item “0” equals the response of “1” on the scale, etc.

Item	Missing data, n (%)	Item frequency distribution, n (%)				
		0	1	2	3	4
1	0 (0)	1 (1)	5 (6)	21 (26)	36 (44)	19 (23)
2	1 (1)	0 (0)	4 (5)	16 (20)	9 (11)	52 (63)
3	0 (0)	1 (1)	0 (0)	10 (12)	24 (29)	47 (58)
4	2 (2)	0 (0)	0 (0)	6 (7)	16 (20)	58 (71)
5	0 (0)	2 (2)	2 (2)	21(26)	32 (39)	25 (31)
6	3 (4)	2 (2)	3 (4)	21 (25)	13 (16)	40 (49)
7	1 (1)	0 (0)	1 (1)	4 (5)	34 (42)	42 (51)
8	2 (2)	1 (1)	1(1)	5 (6)	10 (13)	63 (77)
9	0 (0)	0 (0)	0 (0)	6 (7)	28 (34)	48 (59)
10	0 (0)	0 (0)	1 (1)	4 (5)	14 (17)	63 (77)

Scaling Assumptions

Item means ranged from 2.82 to 3.70 (Table 4). The item-total correlations showed that each item contributed substantially to the total score with correlations ranging from 0.454 to 0.731 (Table 5), thus indicating that the scale can be summarized. The total sum for the SUS in the data ranged between 50 and 100 (mean 84, SD 13).

Table . Item descriptive statistics for the Swedish version of the System Usability Scale (n=81).

Item	Score, mean (SD)	Skewness
1	2.82 (0.904)	−.550
2	3.35 (0.964)	−1.094
3	3.41 (0.800)	−1.490
4	3.65 (0.618)	−1.585
5	2.93 (0.940)	−.766
6	3.09 (1.076)	−.876
7	3.44 (0.652)	−1.038
8	3.66 (0.762)	−2.682
9	3.51 (0.633)	−.943
10	3.70 (0.622)	−2.203

Table . Item-total statistics for the Swedish version of the System Usability Scale (n=81).

Item	Scale mean if item deleted	Corrected item-total correlations	Cronbach α if item deleted
1	30.80	0.463	.847
2	30.30	0.731	.820
3	30.22	0.534	.840
4	30.00	0.657	.833
5	30.72	0.479	.847
6	30.57	0.463	.852
7	30.21	0.454	.846
8	29.99	0.687	.827
9	30.12	0.675	.831
10	29.96	0.585	.837

Acceptability

As indicated by the item mean score (range 2.82-3.70; Table 4) and item-frequency distribution (Table 3), the instrument showed acceptability. For 3 items (4, 8, and 10), the item frequency was above 70%, which is still within the acceptable range for the absence of a ceiling effect. Skewness statistics were below or near the acceptable range of −1 for a total of 7 items, thus indicating that the distribution was excessively skewed (Table 4).

Convergent Validity

Convergent validity was evaluated with the correlation between the total sum of the SUS and the total sum of domain 5 in the eHLQ. As expected, there was a positive correlation between the total score of the instruments (correlation coefficient 0.305), which was significant (*P*=.001) and supported the construct validity of the Swedish version of the SUS.

Internal Consistency and Measurement Error

Cronbach α for the scale was .852. Corrected item-total correlations were between 0.454 and 0.731, as shown in Table 5, indicating internal consistency for the different items. For all items except item 6 (I thought there was too much inconsistency in this system), the α value if the item deleted was lower than the Cronbach α for the scale. The SEM was 5.05 (95% CI –4.84 to 14.954) points for the Swedish version of the SUS.

Discussion

Principal Findings

This study presents a new Swedish version of the SUS that is psychometrically tested. This study seeks to establish the new Swedish version of the SUS as a reliable and valid instrument for assessing system usability. Overall, the psychometric testing showed high data quality, good scaling assumptions, high internal consistency, and fair convergent validity. Together, these analyses support the validity and reliability of the new Swedish version of the SUS.

The translation process of the new Swedish version of the SUS was executed incrementally, both within and outside the research team, involving an authorized translator who conducted a back translation. This method facilitated thorough scrutiny and comparison of the translation from multiple perspectives. The approach also helped reduce the risk of bias and improved the scale's validity by incorporating multiple viewpoints [43].

The psychometric testing regarding internal consistency showed that the Cronbach α values were satisfactory and that all items

contributed to the instrument's total score. This indicates that the Swedish translation of the SUS is a stable instrument to use in a Swedish context.

Convergent validity showed fair correlation between the SUS and domain 5 of the eHLQ, which measured the motivation to engage with digital services (I find that digital technology support me in taking care of my health). This supported our hypothesis that a generally positive attitude toward digital solutions would correlate with a high usability score. This hypothesis is also in line with previous research, which shows that positive expectations directed toward a product generate positive subjective usability ratings [14].

There was an indication of a ceiling effect for 3 items on the Swedish version of the SUS (items 4, 8, and 10). They presented above 70% in item frequency distribution and a high mean total score, which is in line with studies indicating a generally high score for the SUS [10]. Previous studies have also shown ceiling effects of the SUS but only for the total score [15,44], thus the responses on the different items can not be justly compared between studies. This study, however, explores the SUS in much more detail and in line with recommended psychometric evaluations, as it explores the psychometric properties for each item separately [12,37]. Further, the distribution was excessively skewed. This could be an expected result of the nonnormal distribution, since negatively skewed variables are assumed to have a ceiling effect [37,45]. Also, the robustness of skewed distribution as the sole indicator of ceiling effect has been questioned [45].

However, based on the skewness result of the items, the question arises as to whether the range of the response options is wide enough (5-point scale) and if the wording in the response options ("strongly disagree" to "strongly agree") is a sufficient description. It could be interesting to expand the number of response options but, arguably, there could be obstacles with revising such a widely used and widespread instrument.

Regarding the high mean scores for both the total sum and the items of the SUS, it should be noted that the population had a mean age of 33 years and are, therefore, used to digital solutions. This could have influenced their experience of the app. Therefore, the high item scores of the SUS could also indicate a product that is perceived to have usability for this group. A recommendation for future studies would be to test the instrument in other contexts and with different age groups. This app was designed to support users who were in exactly the same situation as the participants in this study: parents in a specific situation. Going home with a child after hospitalization can be stressful; the parents in this study reported this eHealth solution as supportive [46]. The SUS has previously been tested for its sensitivity in different digital solutions and has shown a high scoring (ceiling effect) within the best-of-class products [44].

The Swedish government has envisioned Sweden as a leading country of eHealth in the near future. It has declared the need for individual users to act as the cocreators of such solutions [30]. Within eHealth, there has already been a call for a participatory design to increase the equality of access to digital solutions [47]. Even though digital technology and eHealth are intended to enhance access to health care, for example, despite geographical distances. It may also result in the opposite, as people have different knowledge in using digital solutions. For example, younger people tend to use and access digital solutions differently than older people, which is why solutions should be adapted to the intended end users [48]. In this study, the population was younger, as previously discussed, which could have influenced the result. To have a project design of cocreation often means that end users are involved from the start of the project as collaborators. Coproduction, on the other hand, often involves the end users during the implementation phase [49]. Arguably, the end users are invaluable in all steps of the process, and the usability of products needs to be evaluated where instruments such as the SUS can be useful [50]. This study could therefore be regarded as enhancing the possible participation of end users in the future development of digital solutions and eHealth in a Swedish context.

In conclusion, this study indicates that this developed and psychometrically tested Swedish version of the SUS can be recommended for use within the Swedish adult context.

Limitations

There are some methodological challenges in this study. First, the sample could be regarded as small (N=62), although it was sufficient according to the recommendations [31,33]. A larger sample might enable different methods of analysis, such as Rasch analysis, which might offer deeper psychometric insights.

Second, the 1-year interval for data collection could be considered long in a rapidly changing world. This length of time could influence how the app was perceived. That said, the digital product was not changed to any great extent during this period, so the perception of usability should not have been influenced to any great extent. Additional data collection was carried out after COVID-19 restrictions were essentially withdrawn in Sweden. There is, however, always the possibility of perceptions fluctuating, as the product could be used in various ways according to the circumstances of the parents.

Third, there was an intention to pursue a data quality test for reproducibility as some of the participants answered the SUS twice. Regrettably, the test-retest sample (n=20) was regarded as too small in this study, and such analyses need to be considered in future studies. In addition, the use of imputation can be discussed. There was a low number of missing data, with a noncomplete frequency of only 9 responses in total. The percentage of missing data per item was also low across all items. This resulted in only 6 imputed scores across the items, and based on mean score for the item, this was regarded as not influencing the result noticeably.

Conclusions

This study presents a new Swedish version of the SUS. It is the first study to carry out a psychometric evaluation of a Swedish SUS, to establish the Swedish version of the SUS as a reliable and valid instrument for assessing system usability. Overall, the psychometric testing showed high data quality, good scaling assumptions, high internal consistency, and fair convergent validity, all of which support the validity and reliability of the new Swedish version. The results from this study are promising. They raise the possibility that the Swedish version of the SUS

could be used to evaluate digital health solutions. To further strengthen the usability of the scale, we suggest additional analysis on data that evaluate more complex eHealth technology and include a wider participant age group, both younger and older.

Acknowledgments

We would like to express our sincere gratitude to the participants in the study. We would like to acknowledge Inger Kristensson Hallström, who passed away in 2024, for her extensive work, deep commitment to the field of pediatric nursing, and expertise dedicated to the eChildHealth program. Additionally, we wish to express our appreciation to Magnus C. Persson, for the design of the questionnaires in REDCap (Research Electronic Data Capture) and assisting in data transfer, as well as Desirée Lemmen and other gatekeepers for their assistance with participant inclusion and the distribution of questionnaires to the study participants. This study was funded by the Swedish Research Council for Health, Working Life and Welfare, Sweden (grant 2018-01399).

Authors' Contributions

All authors have made substantial contributions to all of the following: (1) the conception and design of the study, the acquisition of data, or the analysis and interpretation of data; (2) the drafting of this paper or its critical revision for important intellectual content; and (3) final approval of the manuscript version to be submitted.

Conflicts of Interest

None declared.

References

1. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023 Mar 23;10:e45143. [doi: 10.2196/45143] [Medline: 36843321]
2. Lennon MR, Bouamrane MM, Devlin AM, et al. Readiness for delivering digital health at scale: lessons from a longitudinal qualitative evaluation of a national digital health innovation program in the United Kingdom. J Med Internet Res 2017 Feb 16;19(2):e42. [doi: 10.2196/jmir.6900] [Medline: 28209558]
3. Digital health. World Health Organization. 2024. URL: https://www.who.int/westernpacific/health-topics/e-health [accessed 2024-07-01]
4. Brown W III, Yen PY, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087. [doi: 10.1016/j.jbi.2013.08.001]
5. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
6. Dix A. Human-Computer Interaction: Pearson; 2003. URL: https://www.pearson.com/en-gb/subject-catalog/p/human-computer-interaction/P200000008804/9780130461094 [accessed 2025-01-07]
7. Brooke J. SUS: a “quick and dirty” usability. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry: CRC Press; 1996:189-194 URL: https://www.taylorfrancis.com/chapters/edit/10.1201/9781498710411-35/sus-quick-dirty-usability-scale-john-brooke [accessed 2025-01-07]
8. Brooke J. SUS: a retrospective. J User Exp 2013;8(2):29-40 [FREE Full text]
9. Hägglund M, Scandurra I. User evaluation of the Swedish patient accessible electronic health record: System Usability Scale. JMIR Hum Factors 2021 Jul 27;8(3):e24927. [doi: 10.2196/24927] [Medline: 34313596]
10. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum-Comput Interact 2018 Jul 3;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
11. Lewis JR, Sauro J. The factor structure of the System Usability Scale. In: Kurosu M, editor. Human Centered Design. HCD 2009. Lecture Notes in Computer Science, vol 5619: Springer; 2009:94-103. [doi: 10.1007/978-3-642-02806-9_12]
12. Hobart J, Cano S. Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods. Health Technol Assess 2009 Feb;13(12):iii, ix-x, 1-iii, ix-x177. [doi: 10.3310/hta13120] [Medline: 19216837]
13. Wang Y, Lei T, Liu X. Chinese System Usability Scale: translation, revision, psychological measurement. Int J Hum-Comput Interact 2020 Jun 14;36(10):953-963. [doi: 10.1080/10447318.2019.1700644]
14. Raita E, Oulasvirta A. Too good to be bad: favorable product expectations boost subjective usability ratings. Interact Comput 2011 Jul;23(4):363-371. [doi: 10.1016/j.intcom.2011.04.002]
15. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the System Usability Scale. Int J Hum-Comput Interact 2020 Dec 13;36(20):1883-1901. [doi: 10.1080/10447318.2020.1801173]
16. Sharfina Z, Santoso HB. An Indonesian adaptation of the System Usability Scale (SUS). 2016 Presented at: 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS); Malang, Indonesia p. 145-148. [doi: 10.1109/ICACSIS.2016.7872776]

17. Borkowska A, Jach K, Borzemski L, Grzech A, Świątek J, Wilimowska Z, editors. Pre-testing of Polish translation of System Usability Scale (SUS). In: Information Systems Architecture and Technology: Proceedings of the 37th International Conference on Information Systems Architecture and Technology–ISAT 2016–Part I. Advances in Intelligent Systems and Computing, vol 521: Springer; 2017:143-153. [doi: 10.1007/978-3-319-46583-8_12]
18. Streiner DL, Norman GR, Cairney J. Health Measurement Scales: A Practical Guide to Their Development and Use: Oxford University Press; 2024. URL: https://global.oup.com/academic/product/health-measurement-scales-9780199685219?cc=us&lang=en&amp [accessed 2025-01-07]
19. AlGhannam BA, Albustan SA, Al-Hassan AA, Albustan LA. Towards a standard Arabic System Usability Scale: psychometric evaluation using communication disorder app. Int J Hum-Comput Interact 2018 Sep 2;34(9):799-804. [doi: 10.1080/10447318.2017.1388099]
20. Hvidt JCS, Christensen LF, Sibbersen C, Helweg-Jørgensen S, Hansen JP, Lichtenstein MB. Translation and validation of the System Usability Scale in a Danish mental health setting using digital technologies in treatment interventions. Int J Hum-Comput Interact 2020 May 8;36(8):709-716. [doi: 10.1080/10447318.2019.1680922]
21. Ensink CJ, Keijsers NLW, Groen BE. Translation and validation of the System Usability Scale to a Dutch version: D-SUS. Disabil Rehabil 2024 Jan;46(2):395-400. [doi: 10.1080/09638288.2022.2160837] [Medline: 36573399]
22. Moorthy P, Weinert L, Harms BC, Anders C, Siegel F. German version of the mHealth App Usability Questionnaire in a cohort of patients with cancer: translation and validation study. JMIR Hum Factors 2023 Nov 1;10:e51090. [doi: 10.2196/51090] [Medline: 37910144]
23. Borsci S, Federici S, Lauriola M. On the dimensionality of the System Usability Scale: a test of alternative measurement models. Cogn Process 2009 Aug;10(3):193-197. [doi: 10.1007/s10339-009-0268-9] [Medline: 19565283]
24. Mohamad Marzuki MF, Yaacob NA, Yaacob NM. Translation, cross-cultural adaptation, and validation of the Malay version of the System Usability Scale questionnaire for the assessment of mobile apps. JMIR Hum Factors 2018 May 14;5(2):e10308. [doi: 10.2196/10308] [Medline: 29759955]
25. Dianat I, Ghanbari Z, AsghariJafarabadi M. Psychometric properties of the Persian language version of the System Usability Scale. Health Promot Perspect 2014 Jul 12;4(1):82-89. [doi: 10.5681/hpp.2014.011] [Medline: 25097841]
26. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP. European Portuguese validation of the System Usability Scale (SUS). Proc Comput Sci 2015;67:293-300. [doi: 10.1016/j.procs.2015.09.273]
27. Blažica B, Lewis JR. A Slovene translation of the System Usability Scale: the SUS-SI. Int J Hum Comput Interact 2015 Feb;31(2):112-117. [doi: 10.1080/10447318.2014.986634]
28. Sevilla-Gonzalez MDR, Moreno Loaeza L, Lazaro-Carrera LS, et al. Spanish version of the System Usability Scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020 Dec 16;7(4):e21161. [doi: 10.2196/21161] [Medline: 33325828]
29. SUS Svensk: System Usability Scale in Swedish. Rosenfeld Media. 2011 Jul 4. URL: https://rosenfeldmedia.com/sus-svensk-system-usability-scale-in-swedish/ [accessed 2024-05-01]
30. Ministry of Health and Social Affairs, Swedish Association of Local Authorities and Regions. Vision for eHealth 2025 – common starting points for digitisation of social services and health care. e-Hälsa 2025. URL: https://ehalsa2025.se/wp-content/uploads/2021/02/vision-for-ehealth-2025.pdf [accessed 2024-10-01]
31. Mokkink LB, Prinsen C, Patrick DL, et al. COSMIN study design checklist for patient-reported outcome measurement instruments. BMJ. 2019 Jul. URL: https://gut.bmj.com/content/gutjnl/70/1/139/DC1/embed/inline-supplementary-material-1.pdf [accessed 2025-01-07]
32. Lewis JR. Measuring perceived usability: SUS, UMUX, and CSUQ ratings for four everyday products. Int J Hum-Comput Interact 2019 Sep 14;35(15):1404-1419. [doi: 10.1080/10447318.2018.1533152]
33. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine (Phila Pa 1976) 2000 Dec 15;25(24):3186-3191. [doi: 10.1097/00007632-200012150-00014] [Medline: 11124735]
34. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
35. Kristensson Hallström I, Castor C, Hansson H, et al. eHealth as an aid for facilitating and supporting self-management in families with long-term childhood illness; development, evaluation, and implementation in clinical practice – a research programme. Clin Health Promot 2018;13(1):e23002 [FREE Full text]
36. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth Literacy Framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
37. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome. Health Technol Assess 2004 Mar;8(9):iii, 1-iii,48. [doi: 10.3310/hta8090] [Medline: 14982653]
38. Huisman M. Imputation of missing item responses: some simple techniques. Qual Quant 2000;34(4):331-351. [doi: 10.1023/A:1004782230065]

39. Sjöström AE, Hajdarevic S, Hörnsten Å, Kristjánsdóttir Ó, Castor C, Isaksson U. The Swedish version of the eHealth Literacy Questionnaire: translation, cultural adaptation, and validation study. J Med Internet Res 2023 Apr 12;25(1):e43267. [doi: 10.2196/43267] [Medline: 37043268]
40. Abma IL, Rovers M, van der Wees PJ. Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. BMC Res Notes 2016 Apr 19;9:226. [doi: 10.1186/s13104-016-2034-2] [Medline: 27094345]
41. Samuelsson M, Wennick A, Bengtsson M, Lydrup ML, Jakobsson J. Translation, cultural adaptation, and psychometric testing of the Supportive Care Needs Survey for partners and caregivers for Swedish family members of persons diagnosed with colorectal cancer. J Patient Rep Outcomes 2023 Oct 11;7(1):100. [doi: 10.1186/s41687-023-00636-1] [Medline: 37819416]
42. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human participants. JAMA 2024 Oct 19. [doi: 10.1001/jama.2024.21972] [Medline: 39425955]
43. Polit DF, Beck CT. Nursing Research: Generating and Assessing Evidence for Nursing Practice, th edition: Wolters Kluwer; 2021.
44. Kortum PT, Bangor A. Usability ratings for everyday products measured with the System Usability Scale. Int J Hum Comput Interact 2013 Jan;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
45. Ho AD, Yu CC. Descriptive statistics for modern test score distributions: skewness, kurtosis, discreteness, and ceiling effects. Educ Psychol Meas 2015 Jun;75(3):365-388. [doi: 10.1177/0013164414548576] [Medline: 29795825]
46. Lindkvist RM, Sjöström-Strand A, Landgren K, Johnsson BA, Stenström P, Hallström IK. "In A Way We Took the Hospital Home"-a descriptive mixed-methods study of parents' usage and experiences of eHealth for self-management after hospital discharge due to pediatric surgery or preterm birth. Int J Environ Res Public Health 2021 Jun 15;18(12):6480. [doi: 10.3390/ijerph18126480] [Medline: 34203985]
47. Kirklies PC, Neumann O, Hohensinn L. Promoting digital equality in co-production: the role of platform design. Gov Inf Q 2024 Mar;41(1):101903. [doi: 10.1016/j.giq.2023.101903]
48. Frennert S. Hitting a moving target: digital transformation and welfare technology in Swedish municipal eldercare. Disabil Rehabil Assist Technol 2021 Jan;16(1):103-111. [doi: 10.1080/17483107.2019.1642393] [Medline: 31348681]
49. Lember V, Brandsen T, Tõnurist P. The potential impacts of digital technologies on co-production and co-creation. Pub Manag Rev 2019 Nov 2;21(11):1665-1686. [doi: 10.1080/14719037.2019.1619807]
50. van Velsen L, Evers M, Bara CD, Op den Akker H, Boerema S, Hermens H. Understanding the acceptance of an eHealth technology in the early stages of development: an end-user walkthrough approach and two case studies. JMIR Form Res 2018 Jun 15;2(1):e10474. [doi: 10.2196/10474] [Medline: 30684434]

Abbreviations

COSMIN: Consensus-Based Standards for the Selection of Health Status Measurement Instruments
eHLQ: eHealth Literacy Questionnaire
SUS: System Usability Scale
WHO: World Health Organization

Edited by A Kushniruk, E Borycki; submitted 13.07.24; peer-reviewed by JS Berner, J Brooke; revised version received 04.12.24; accepted 08.12.24; published 16.01.25.

Please cite as:
Åvik Persson H, Castor C, Andersson N, Hylén M
Swedish Version of the System Usability Scale: Translation, Adaption, and Psychometric Evaluation
JMIR Hum Factors 2025;12:e64210
URL: https://humanfactors.jmir.org/2025/1/e64210
doi:10.2196/64210

Original Paper

Evaluating the Development, Reliability, and Validation of the Tele-Primary Care Oral Health Clinical Information System Questionnaire: Cross-Sectional Questionnaire Study

Rosnah Sutan[1*], MPH, MD, PhD; Shahida Ismail[1*], MPH, MD; Roszita Ibrahim[1*], BSc, MSc, PhD

Department of Public Health Medicine, Faculty of Medicine, Universiti Kebangsaan Malaysia, Cheras Kuala Lumpur, Malaysia

[*]all authors contributed equally

Corresponding Author:
Rosnah Sutan, MPH, MD, PhD
Department of Public Health Medicine
Faculty of Medicine
Universiti Kebangsaan Malaysia
Jalan Yaakob Latiff, Bandar Tun Razak Cheras
Cheras Kuala Lumpur, 56000
Malaysia
Phone: 60 391459587
Email: rosnah.sutan@hctm.ukm.edu.my

Abstract

Background: Evaluating digital health service delivery in primary health care requires a validated questionnaire to comprehensively assess users' ability to implement tasks customized to the program's needs.

Objective: This study aimed to develop, test the reliability of, and validate the Tele-Primary Care Oral Health Clinical Information System (TPC-OHCIS) questionnaire for evaluating the implementation of maternal and child digital health information systems.

Methods: A cross-sectional study was conducted in 2 phases. The first phase focused on content item development and was validated by a group of 10 experts using the content validity index. The second phase was to assess its psychometric testing for reliability and validity.

Results: A structured questionnaire of 65 items was constructed to assess the TPC-OHCIS delivery for primary health care use based on literature and has been validated by 10 experts, and 319 respondents answered the 65-item TPC-OHCIS questionnaire, with mean item scores ranging from 1.99 (SD 0.67) to 2.85 (SD 1.019). The content validity, reliability, and face validity showed a scale-level content validity index of 0.90, scale-level content validation ratio of 0.90, and item-level face validity index of 0.76, respectively. The internal reliability was calculated as a Cronbach α value of 0.90, with an intraclass correlation coefficient of 0.91. Scales were determined by the scree plot with eigenvalues >1, and 13 subscales were identified based on principal component analysis. The Kaiser-Meyer-Olkin value was 0.90 (P<.049). The total variance explained was 76.07%, and factor loading scores for all variables were >0.7. The Bartlett test of sphericity, determining construct validity, was found to be significant (P<.049).

Conclusions: The TPC-OHCIS questionnaire is valid to be used at the primary health care level to evaluate the TPC-OHCIS implementation. It can assess health care workers' work performance and job acceptance and improve the quality of care.

(JMIR Hum Factors 2025;12:e53630) doi:10.2196/53630

KEYWORDS

telehealth; electronic health; eHealth; public health information system; psychometric analysis

Introduction

Background

Digital health is mostly used interchangeably with eHealth, telehealth, or mobile health in the literature [1,2]. It requires integrated and interdisciplinary sector involvement to use knowledge information and communication technology in health (eg, medicine, public health, pharmaceutical, dentistry, and health management) [1-8]. It enables the national health system to ensure population access to health services and the ability to monitor and evaluate health system delivery performance [1,2,9]. Electronic health records are created from digital health systems for case management monitoring and can be shared

across health care settings [9]. The data sharing platform allows access to integration interfaces that include electronic medical records, appointments, electronic prescriptions, e-commerce, public health surveillance, system monitoring such as vaccination, environmental health, institutional health management, and an online platform for teaching and learning among health care workers (HCWs) [10-16].

Malaysian Digital Health

Malaysia began integrating digital health into its health care system in 1998 [4,5]. The Telemedicine Act 1997 (Act 564) was enacted to regulate telemedicine practices in the country [4,5,17-25]. Later, the telemedicine blueprint was created to outline the government's vision for digital health implementation and align it with the 7 National Multimedia Super Corridor flagship applications [19-22]. The initiative was referred to as telemedicine and later restated as telehealth. The telehealth system serves as a platform for digital health services in Malaysia [20]. The National Telehealth Policy was launched to support the Vision 2020 agenda, focusing on four key components: (1) lifetime health plan, (2) mass customized health information and education, (3) continuous medical education, and (4) teleconsultation application [1]. The policy was formulated during the Eighth Malaysia Plan and managed by the National Health Informatics Centre Division [4,17,19-22]. The lifetime health plan covers health services from womb to tomb [19]. The telehealth system also includes the clinical support system and health information management and support services, and it encompasses the hospital information system and the Tele-Primary Care Oral Health Clinical Information System (TPC-OHCIS) [18-24]. The TPC-OHCIS is a comprehensive electronic medical record system for primary health care [22-24]. It was initially developed for outpatient services at primary health care clinics (PHCs), and later, it incorporated maternal and child health (MCH), oral health, and other life stage health services to ensure continuity of care [21,22]. The development of TPC-OHCIS was a collaborative effort between the Ministry of Science, Technology, and Innovation; the Ministry of Health, Malaysia; and the Malaysian Institute of Microelectronic Systems, Berhad [21-24]. The system was first tested in the PHCs of Seremban District, Negeri Sembilan, and later expanded to 34 PHCs across 3 states (ie, Perlis, Sarawak, and Selangor) [17,21]. To date, the TPC-OHCIS has been implemented in 108 PHCs across 7 additional states [22,25]. The TPC-OHCIS is a web-based platform that allows HCWs to enter data during clinic services or home visits, automatically updating when connected to the internet [23,24].

Role of Digital Health in Service Quality Performance

The recent COVID-19 pandemic revealed a significant public health issue in health care system delivery to provide comprehensive quality care [2,26-28]. Malaysia has experienced various health service delivery disruptions at PHCs during the critical phase of the COVID-19 pandemic [29]. MCH services include a wide range of services covering school health programs that experienced substantial disruptions due to closing and movement control orders, thus preventing mothers and children from receiving adequate health care services [22]. During the COVID-19 pandemic, many countries have improved health care delivery through digital technology, enhanced resource coordination, and facilitated universal health coverage [29,30].

The HCWs at the PHCs worked on the frontlines, assessing risks, monitoring care treatment, and promoting health empowerment in the community [17]. Most of the administration work related to patient care is recorded manually. Even with the implementation of TPC-OHCIS, many facilities still need to record data manually and enter it into the system because of internet instability at some PHCs [22]. At present, there is no specific policy published by the Ministry of Health to completely replace manual recording of patient information monitoring with digital health. The HCWs monitor patients during home visits and conduct outreach activities to cover areas inaccessible to health facilities [17,22]. Therefore, it is important to continue patient using manual data recording when services are provided offline. The TPC-OHCIS is an electronic medical record system used as a daily operating system and in "real time" at PHCs [22-24]. However, the TPC-OHCIS implemented in selected facilities in Malaysian PHCs is used mostly to record data and monitor patient care only [22-24].

Specific Study Measurement Tool

The conceptualization of this study diverges from earlier studies primarily in its integration of advanced technological modalities and emphasis on patient-centered care delivery, unlike traditional telehealth frameworks, which often focus on providing remote consultations or basic medical services [1,3,18,25]. The TPC-OHCIS incorporates elements of comprehensive primary care delivery, leveraging telehealth technologies to facilitate longitudinal patient-provider relationships, care coordination, and proactive health management. Moreover, this study's conceptualization places a heightened emphasis on the integration of patient health data, wearable devices, and digital health platforms to enhance care delivery and patient engagement. The TPC-OHCIS can remotely monitor patient health metrics, deliver personalized interventions, and empower individuals to take an active role in their health management [24]. The importance of interdisciplinary collaboration and team-based care is crucial in the implementation of the TPC-OHCIS at PHCs. Hence, it is important to assess the implementation of TPC-OHCIS by focusing on technology, organization, environment, or human resource components, as suggested in the literature [3-5,9]. Earlier studies were conducted to evaluate the effectiveness of the TPC-OHCIS in improving health care quality; however, there is limited evidence on assessing HCWs' perceived usefulness and ease of use [9,18,25]. A comprehensive study of HCWs' perspectives is important to provide a shred of extensive knowledge and evidence on the TPC-OHCIS implementation in PHCs.

Therefore, this study aimed at developing, testing reliability, and validating the TPC-OHCIS questionnaire, which was designed as a survey instrument to collect data on the implementation of the TPC-OHCIS at the PHC setting among HCWs related to various aspects of primary care services, including MCH services. It aims to ensure that the TPC-OHCIS questionnaire is a robust and effective tool for assessing the

implementation of MCH services, which are the core services for PHCs. The research question formulated was as follows: "Is a customized questionnaire on digital health information system (TPC-OHCIS) able to assess the HCWs' perception of its implementation process in the delivery of MCH services at the primary care level?" A thorough design of the TPC-OHCIS questionnaire validation helps increase the questionnaire's relevance and usefulness for decision-making purposes, as it is designed to facilitate monitoring and evaluating the TPC-OHCIS operability among the HCWs working at PHCs. To the best of our knowledge, there is no valid questionnaire available to measure the implementation evaluation of any digital health information system for the primary care level that focuses on MCH services, which is a priority service component of PHC and which was monitored regularly for sustainable development goals performance achievement [17].

Methods

Study Design

This study involved several steps for questionnaire development (phase 1), reliability, and validation (phase 2).

Phase 1: Questionnaire Development

The questionnaire is adapted from literature and document review [4,5,9,17-25,31-36]. The TPC-OHCIS questionnaire was developed based on various theoretical models, which may address multiple aspects of remote primary care and patient information systems. The questionnaire was created based on a combination of the technology-organization-environment (TOE) theory [31], the technology acceptance model (TAM) theory [32,33], the human organization technology-fit (HOT-fit) model [34,35], and the diffusion of innovation (DOI) theory [36]. There was a 65-item questionnaire with a 4-point Likert scale (1=highly disagree, 2=disagree, 3=agree, and 4=highly agree) developed based on the aforementioned theories [31-36]. The score scale is created according to the Likert scale that indicates the following: 1=strongly agree, 2=agree, 3=disagree, and 4=strongly disagree. In this study, the researcher did not put a neutral on a scale of 3 to avoid respondent bias [37]. The 4-point Likert scale does not impact the reliability and validity of the questionnaire [2,37].

The development of the questionnaire items was partly adapted from various questionnaires available from previous studies conducted in Malaysia, based on selected theoretical models, using a 4-point scale questionnaire: (1) technology, (2) organization, (3) environment, and (4) human [31-36] (Multimedia Appendix 1 [4,5,9,17-25,31-36]). Therefore, we classified the 65-item TPC-OHCIS questionnaire into 4 scales (Multimedia Appendix 2), described below.

- Domain A, technology: this comprised 17 items and four subscales that include (1) relative advantage (items 1-5), (2) compatibility (items 6-9), (3) complexity (items 10-13), and (4) security concern (items 14-17).
- Domain B, organization: this contained 18 items and four subscales that include (1) the presence of a specified liaison officer (BCHAMP: items 18-22), (2) infrastructure (BINFRA: items 23-26), (3) top management support (BTP: items 27-31), and (4) financial resources (BFIN: items 32-35).
- Domain C, environment: this focuses on vendor support (CVEN: items 36-39).
- Domain D, human: this contained 26 items and six subscales that include (1) staff competency in information systems (DPT: items 40-44), (2) knowledge of the TPC-OHCIS system (DEISK: items 45-50), (3) clinical information technology competency (DCIT: items 51-54), (4) perceived innovativeness of the IT officer (DCIO: items 55-57), (5) perceived ease use (DPEU: 58-61), and (6) perceived usefulness (DPU: items 62-65).

The 65-item TPC-OHCIS questionnaire was developed based on the requirement of the TPC-OHCIS implementation plan for the PHCs in Malaysia. A total of 15 items (item numbers 10, 11, 12, 13, 33, 40, 41, 42, 48, 49, 50, 53, 54, 58, and 39) were written as negative items. The Likert scale score of negative items was reversed for scoring analysis before data were entered into the SPSS software (version 26.0; IBM Corp).

Phase 2: Questionnaire Validation

On the basis of the steps suggested by Boateng et al [38], the second phase involved scale development, which consists of pretesting questions, sampling and survey administration, item reduction, and extraction of latent factors. The scale evaluation requires tests of dimensionality, reliability, and validity. A cross-sectional study was conducted to assess the psychometric properties of the questionnaire.

Study Sample

The sample for each step was calculated and listed in Table 1.

The sample-to-item ratio was decided based on the number of items in the questionnaire. The ratio per item was determined using 5 to 20 samples per item [39,40]. In this study, the survey included 65 items. Therefore, the total calculated sample size required was 325 respondents.

Table 1. Calculated sample size for each validation step.

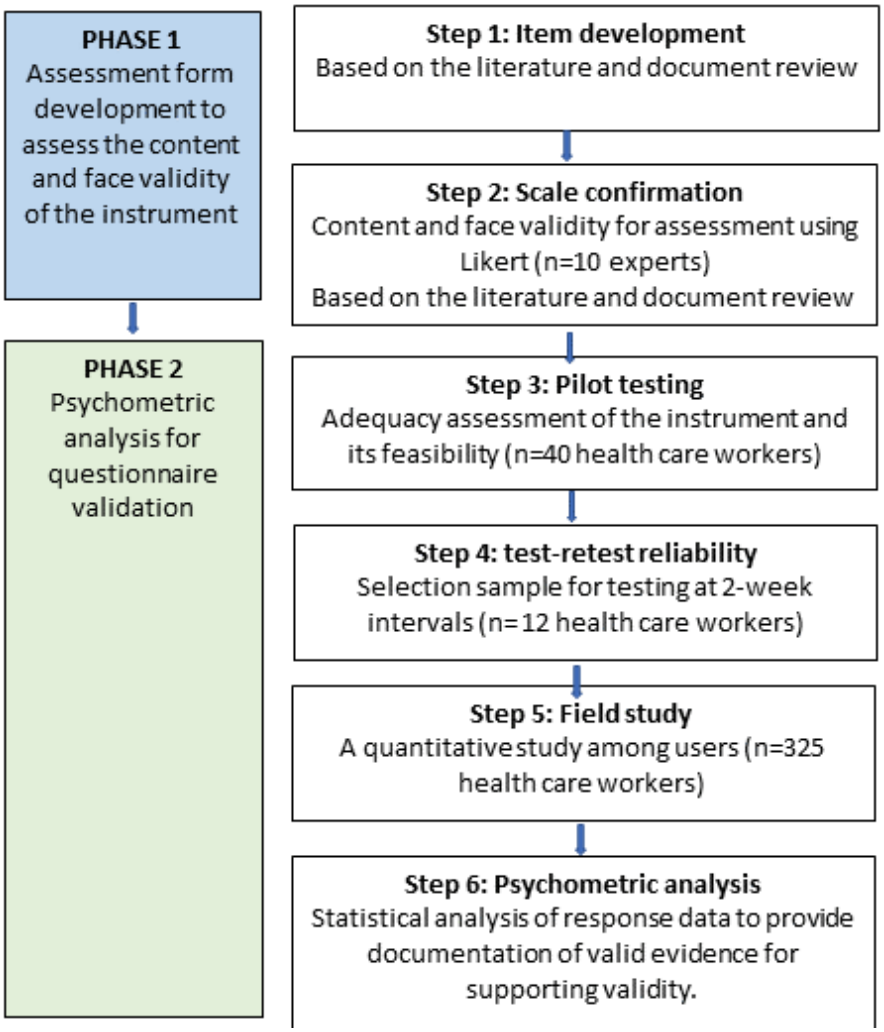
Steps and aims	Sample size required, n	Reference
Pilot study and face validity: to test the adequacy of instrumentation in which the outcome is in the form of a scale	40	[3]
Test-retest reliability: to test the degree of consistency exhibited when a measurement is repeated under identical conditions	12	[4]
Field survey: to test the psychometric analysis of the questionnaire based on tests of dimensionality, reliability, and validity	325	[5]

Content Validity Index

The flow process of questionnaire validation was conducted based on Figure 1.

On the basis of Figure 1, a total of 6 steps were taken to perform content validity for the questionnaire.

Figure 1. Flow process for questionnaire validation.



Step 1: Assessment Form Development for Content and Face Validity

In our literature review search [37,39-41], we listed items to determine the content and face validity. We used the clarity scale of 1 to 3 to indicate the relevance of each item (1=not clear, 2=revision needed, and 3=very clear). The essentiality of the questionnaire was identified using a 1 to 3 scale (1=not essential, 2=useful but not essential, and 3=essential). A similar form was used to assess face validity in evaluating the language, feasibility, readability, and style formatting consistency [37,41].

Step 2: Content and Face Validity (Expert Review)

Face validity was conducted among the experts who were involved in designing and using the system at the primary care level. Ten experts were invited to review the 65-item Tele-Primary Care Oral Health Clinical Information System (TPC-OHCIS) questionnaire. All experts have extensive experience and knowledge related to the TPC-OHCIS. Of the 10 experts, 3 (30%) were from the TPC-OHCIS division, 3 (30%) were public health researchers, 2 (20%) were from the maternal and child health unit services in the Selangor State Health Department, and finally, 2 (20%) were individuals who worked directly with the TPC-OHCIS system and also served as system liaison officers at health clinics. The adaptation of questionnaires from earlier studies [41-43] was checked for content suitability and appropriateness before forward translation was performed. Scale evaluation was performed by experts using a developed form. Owing to time constraints and lockdown because of the COVID-19 pandemic, a face-to-face method for collecting data from the experts was impossible. A Google form (Google LLC) link was created and sent via WhatsApp (Meta Platforms, Inc) to assess the face and content validity. The selected 10 experts were asked to evaluate the scale for content and face validation. All 10 experts' comments were used to improve the questionnaire. Experts were encouraged to provide both verbal and written comments regarding the questionnaire items. Verbal comments were collected via phone call, and written comments were collected via the Google form.

Step 3: Pilot Study

A pilot study was conducted to assess the adequacy of the instrumentation, which measured outcomes to ensure that the instruments used were appropriate and effective for the intended purpose. Through the pilot study, face validity was evaluated to confirm that the 65-item TPC-OHCIS questionnaire developed was constructed accurately and comprehensively. We purposively invited health care workers (HCWs) who were involved in the initial pilot-tested version of the TPC-OHCIS in Seremban District, Negeri Sembilan [17]. The questionnaire was sent after a thorough face and content validity assessment by the experts. Correction and fine-tuning were conducted by the researchers based on their feedback.

Step 4: Test-Retest Reliability

We selected 12 HCWs who were familiar with the TPC-OHCIS involved in the pilot study mentioned above to test the degree of consistency exhibited when a measurement is repeated under identical conditions. The same questionnaire was sent to the same respondents 2 weeks after the first time the survey was conducted.

Step 5: Field Study

A cross-sectional study was performed to assess the item reduction and extraction factors among the HCWs of the TPC-OHCIS public health care clinics (PHCs) in selected states in Malaysia. The TPC-OHCIS is a system used in the PHCs. The PHC is headed by a family medicine specialist who supervises staff from various disciplines (eg, physicians, nurses, medical assistants, pharmacists, laboratory technicians, and record officers).

Step 6: Performing an Analysis for Content Validity, Face Validity, and Test-Retest Reliability

The item-level content validity index (I-CVI) and scale-level content validity index (S-CVI) were used. I-CVI is defined as the proportion of content experts giving items a relevance rating of 3 and 4. S-CVI was calculated by the proportion of understanding by experts (universal agreement [UA] of S-CVI), which gave ratings of 3 and 4, or by average scores given (average of S-CVI). The average of S-CVI is defined as the average of the I-CVI scores for all items on the scale or the average of proportion relevance judged by all experts. The UA of S-CVI is the proportion of items on the scale that achieve a relevance scale of 3 or 4 by all experts. UA score is given as 1 when the item achieved 100% agreement by all experts; otherwise, the UA score is given as 0 [39]. First, we calculated the experts in agreement. Second, a score of 1 was given to items that achieved 100% agreement by all experts, while a score of 0 was given to items that did not achieve 100% agreement. Third, experts in the agreement were divided by the total number of experts (calculated for I-CVI). The fourth step was to divide the results obtained in the third step by 65, which was the total number of items in the questionnaire (calculated for average of S-CVI). Finally, we divided the UA result from step 2 by 65 items (calculated for UA of S-CVI). The face validity index was used to evaluate the form of clarity and comprehensibility of language and instructions used in the questionnaire [41]. The respondents were requested to rate the comprehensibility of each item using a scale of 1 to 4 (1=not understandable, 2=somewhat understandable, 3=understandable, and 4=very understandable). The item-level face validity index was computed for each item by dichotomizing the 4-point scale, where items scoring either 1 or 2 were recorded as 0 and items scoring 3 and 4 were recorded as 1. The values later were added up according to each item, and the total values were divided by the total number of respondents. Test-retest reliability was used to define whether the questionnaire was answered by respondents due to chance. The internal reliability of questionnaires was assessed using the Cronbach α coefficient. The Pearson correlation and intraclass correlation coefficients of the scores of the 2 tests were calculated. The test-retest reliability is achieved when the value of intraclass correlations is 0.6 to 0.8 (good reliability), and values >0.8 indicate excellent reliability, which means the higher the correlation, the higher the test-retest reliability, with values close to 0 indicating low reliability [38].

Forward-Backward Translation

The questionnaire was translated from English to the Malay language, then underwent content assessment, and its cross-cultural validity was evaluated by 3 experts representing Malaysia's diverse ethnic backgrounds (Malay, Chinese, and Indian), proficient in both English and Malay. After the questionnaire had been reviewed, the forward-backward translation was sent to Proofreaders United company for translation and proofreading.

Ethical Considerations

Ethics approval was granted by the National University of Malaysia (FF-2021-124) and the Ministry of Health Malaysia National Medical Research Registry (NMRR-21-599-58521, investigatory initiated research). Meanwhile, the researcher also needed to obtain verbal permission and a signed consent letter from the respondent as evidence of consent to participate in the study. Respondents were informed that their participation was voluntary and they would not receive any payment. All research information collected was treated as confidential. The data are displayed anonymously without the name, address, or any identity that describes the respondent when presented as the study output.

Statistical Analysis

The internal reliability was assessed using Cronbach α. The validity index was calculated based on the content validity index (CVI). The Bartlett test of sphericity was used to determine the construct validity. The Kaiser-Meyer-Olkin measure and total variance explained (TVE) were evaluated. Factor analysis was performed using varimax rotation in principal component analysis (PCA) to verify construct validity (ie, discriminant and convergent validity). Items loading >0.60 were considered for further analysis [38].

Results

CVI Evaluation

Out of 10 experts evaluating the 65-item TPC-OHCIS questionnaire (TPC-OHCIS questionnaire in Multimedia

Appendix 2), the scale-level CVI (S-CVI) average for item 1 of domain A, relative advantage, to item 65 of domain D, perceived usefulness, was 0.9. The average of S-CVI (based on item-level CVI) was 7.415, the average of S-CVI (based on proportion) was 0.742, and the universal agreement of S-CVI was 0.045. The universal agreement of S-CVI was calculated based on the proportion of items on the scale that obtained a relevance value of 3 or 4 from all the experts (Multimedia Appendix 3).

Face Validation

The face validity index was assessed for clarity, language comprehensibility, and instructions used in the questionnaire. The calculated item face validity index was 0.785 and fulfilled the face validity criteria, which was >0.6 [41].

Test-Retest Reliability

The 65-item survey's internal reliability score for Cronbach α was 0.90 for the whole instrument and exceeded the suggested minimum value of Cronbach α=0.70 [38]. Pearson correlation coefficient was 0.90, and the intraclass correlation coefficient was 0.91.

Descriptive Analysis

There were 319 respondents (Table 2) who answered the 65-item TPC-OHCIS questionnaire, with mean item scores ranging from 1.99 to 2.85 (Table 3). Our study found that a high mean score was achieved in all subscales except for the item 17 in domain A, security concerns (*how much can I trust the vendor)*. The subscale mean scores were lowest at 1.93 (SD 0.751) and 1.99 (SD 0.832) for item 63 in domain D, perceived usefulness (*TPC-OHCIS application improved the quality of my work, respectively).* The mean score for communication was the highest for item 38 in domain C, vendor, with a mean score of 3.45 (SD 0.819), and item 39 in domain C, vendor, with a mean score of 3.39 (SD 0.805).

Table 2. Respondent profile for field study (N=319).

Variables and category	Frequency, n (%)
Occupation	
Family medicine specialist	5 (1.6)
Medical officer	69 (21.6)
Matron	2 (0.6)
Head nurse	17 (5.3)
Nurse	138 (43.3)
Community nurse	88 (27.6)
Sex	
Male	6 (1.9)
Female	313 (98.1)
Race	
Chinese	7 (2.2)
Indian	21 (6.6)
Malay	285 (89)
Others	6 (1.9)

Table 3. Descriptive statistics for item mean score of the Tele-Primary Care Oral Health Clinical Information System (TPC-OHCIS) questionnaire.

Item question	Question statement	Scores, mean (SD)
ARA[a] 1	Using the TPC-OHCIS application enables me to do my work quickly.	2.05 (0.865)
ARA 2	Using the TPC-OHCIS application improves the quality of my work.	1.99 (0.832)
ARA 3	Using the TPC-OHCIS application enhances my effectiveness on the job.	2.01 (0.864)
ARA 4	Using the TPC-OHCIS application increases my productivity.	2.04 (0.825)
ARA 5	Using the TPC-OHCIS application makes my job easier.	2.02 (0.795)
ACOM[b] 6	TPC-OHCIS application can be easily accessed across multiple platforms (laboratory results, x-ray, and other related patient data).	2.53 (0.941)
ACOM 7	TPC-OHCIS user interfaces provide transparent access to all platforms (e-notification and VEKPRO[c]).	2.62 (0.931)
ACOM 8	Data received from other devices (tablet/laptop/ smartphone) outside health facilities in the TPC-OHCIS application can be easily merged into the database for analysis.	2.74 (0.932)
ACOM 9	Information is shared seamlessly across our organization regardless of location.	2.74 (0.839)
ACOMPLEX[d] 10	I do not know enough about the TPC-OHCIS application to handle my job satisfactorily.	2.86 (0.870)
ACOMPLEX 11	I need a long time to understand and get familiar with the TPC-OHCIS application.	2.82 (0.961)
ACOMPLEX 12	I do not find enough time to study and upgrade my technology skills before using the TPC-OHCIS system.	2.71 (0.870)
ACOMPLEX 13	I often find it too complex for me to understand and use the TPC-OHCIS application.	2.99 (1.019)
ASEC[e] 14	I feel secure in using the TPC-OHCIS application, keying in patients' data, and sharing it across my organization.	2.05 (0.855)
ASEC 15	I would feel safe using the TPC-OHCIS application to retrieve patient data.	2.80 (0.754)
ASEC 16	I am concerned about data patient leakage.	2.07 (0.785)
ASEC 17	I am concerned about how much I can trust the vendor.	1.93 (0.751)
BCHAMP[f] 18	A specified liaison officer will provide useful information to top managers and vendors about the TPC-OHCIS application faulty.	2.02 (0.736)
BCHAMP 19	A specified liaison officer plays a role in upgrading the TPC-OHCIS application for users' needs.	2.01 (0.760)
BCHAMP 20	A specified liaison officer has a good relationship with both vendors and top managers.	2.22 (0.851)
BCHAMP 21	A specified liaison officer can bring staff to use the TPC-OHCIS application well.	2.41 (0.972)
BCHAMP 22	A specified liaison officer provides training/courses for the users a few times a year.	2.54 (0.895)
BINFRA[g] 23	We have enough computers for the TPC-OHCIS application use.	2.27 (0.743)
BINFRA 24	We have a reliable computer network in our use.	2.11 (0.767)
BINFRA 25	Appropriate hardware, software, and network infrastructure were in place before TPC-OHCIS implementation.	1.90 (0.749)
BINFRA 26	We have integrated IS[h] applications encompassing different functional areas (laboratory and pharmacy).	2.05 (0.775)
BTP[i] 27	Top management always supports and encourages the use of the TPC-OHCIS application for job-related tasks.	2.15 (0.785)
BTP 28	Top management provides most of the necessary help and resources to enable people to use the TPC-OHCIS system.	2.17 (0.746)
BTP 29	Top management provides good access to hardware when staff need it.	2.15 (0.732)
BTP 30	Top management gives feedback to vendors on every dismayed or unsatisfactory comment from staff.	2.44 (0.822)

Item question	Question statement	Scores, mean (SD)
BTP 31	Top management provides good access to the TPC-OHCIS application when staff need it.	2.55 (0.842)
BFIN[j] 32	There is enough financial aid from the organization for the coordination of the system implementation.	2.44 (0.901)
BFIN 33	I find difficulties in using the TPC-OHCIS application because it cannot be upgraded due to not having enough budget.	2.75 (0.814)
BFIN 34	Enough computers are available to access the TPC-OHCIS application.	2.29 (0.743)
BFIN 35	We easily obtain obsolete computer replacements.	2.28 (0.729)
CVEN[k] 36	Vendors entertain each of our complaints dutifully.	2.41 (0.704)
CVEN 37	The vendor can upgrade the TPC-OHCIS according to our needs.	2.41 (0.771)
CVEN 38	The system vendor attends technical meetings quite frequently.	3.45 (0.819)
CVEN 39	I have a platform to voice out problems regarding the TPC-OHCIS application directly to the vendor.	3.39 (0.805)
DPTC[l] 40	I do not know how to use computers.	3.19 (0.874)
DPTC 41	I never used to work online.	2.11 (0.849)
DPTC 42	I need people's help to use a computer.	2.15 (0.859)
DPTC 43	I like to work using the online system.	2.29 (0.836)
DPTC 44	The TPC-OHCIS application is easy to operate.	2.33 (0.794)
DEISK[m] 45	I have enough training before working with the TPC-OHCIS application.	2.25 (0.836)
DEISK 46	It took me only a few days before I could master the TPC-OHCIS application well.	2.85 (0.920)
DEISK 47	The TPC-OHCIS facilitates task management.	2.63 (0.912)
DEISK 48	The TPC-OHCIS application is hard to use.	2.61 (0.900)
DEISK 49	I have to open many interfaces just to key in 1 patient's data.	2.05 (0.730)
DEISK 50	The TPC-OHCIS application takes much time because I have to open so many interfaces.	2.50 (0.715)
DCIT[n] 51	I have confidence in my ability to operate the TPC-OHCIS application.	2.60 (0.725)
DCIT 52	I have the expertise regarding Information technology to provide valuable knowledge to the organization.	2.23 (0.737)
DCIT 53	It does not make any difference whether I add/share knowledge with others related to the use of the TPC-OHCIS application.	2.22 (0.688)
DCIT 54	I feel that other employees can provide more valuable knowledge about the system's use.	2.22 (0.688)
DCIO[o] 55	The ITO[p] is actively considering the introduction of new technology to solve to organization's problem.	2.34 (0.672)
DCIO 56	The ITO tries to keep a technological leading edge by adopting new technology early.	2.84 (0.823)
DCIO 57	The ITO tends to take risks in the decision-making of new technology introduction.	2.80 (0.847)
DPEU[q] 58	I often become confused every time I use the TPC-OHCIS application.	2.11 (0.810)
DPEU 59	Interacting with the TPC-OHCIS application is frequently frustrating.	2.18 (0.768)
DPEU 60	I find that the TPC-OHCIS application makes my job easier.	2.28 (0.846)
DPEU 61	The TPC-OHCIS application provides useful guidance in performing tasks.	2.06 (0.764)
DPU[r] 62	My job would be hard to perform without the TPC-OHCIS application.	2.13 (0.835)
DPU 63	Using the TPC-OHCIS application improves my job performance.	2.18 (0.767)
DPU 64	Using the TPC-OHCIS application saves me time.	2.13 (0.851)

Item question	Question statement	Scores, mean (SD)
DPU 65	Using the TPC-OHCIS application supports critical aspects of my job (e.g, retrieving patients with missed treatment).	2.17 (0.782)

[a]ARA: domain A, relative advantage.
[b]ACOM: domain A, compatibility.
[c]VEKPRO: vector program for reporting vector-borne diseases and outbreaks.
[d]ACOMPLEX: domain A, complexity.
[e]ASEC: domain A, security.
[f]BICHAMP: domain B, champion.
[g]BINFRA: domain B, infrastructure.
[h]IS: information system.
[i]BTP: domain B, top management support.
[j]BFIN: domain B, financial support.
[k]CVEN: domain C, vendor.
[l]DPTC: domain D, perceived technical competence.
[m]DEISK: domain D, employee information system knowledge.
[n]DCIT: domain D, competency of employee's IT.
[o]DCIO: domain D, chief information officer innovativeness.
[p]ITO: IT officer.
[q]DPEU: domain D, perceived ease of use.
[r]DPU: domain D, perceived usefulness.

Statistical Analysis

The 65-item TPC-OHCIS questionnaire was created based on a mix of items adapted from various sources in the development of the questionnaire. It was adapted by a combination of the TOE theory [31], the HOT-fit model [34,35], the DOI theory [36], and the TAM theory [32,33]. The content validity was assessed by experts according to the 4 scales (ie, technology, organization, environment, and human), as highlighted in the previous studies [33-35]. In this study, we explored the constructs generated from this initial group of 65 items, which were developed based on their multidimensional nature. An exploration of the subdimensions based on the 65-item TPC-OHCIS questionnaire at the initial phase found that there were 13 factors, without being restricted to 4 domains. Exploratory factor analysis (EFA) determined the quantity of components (or themes) that emerged for the questionnaire items. The process involved grouping the measurements of comparable themes. The scree plot reveals that there were 13 components in a 65-item questionnaire (Figure 2).

A PCA with varimax rotation yielded 13 components with an eigenvalue of >1. The eigenvalue for the first component was 17.71 and accounted for 27% of the variance. The difference between the first and second components was 11.98, while the subsequent eigenvalues were small (4.91, 4.64, and 3.40). The Kaiser-Meyer-Olkin value was 0.908, which showed sample adequacy. The Bartlett test of sphericity value for the approximate chi-square was χ^2_{2080}=18,219.9; this was significant (P=.001).

Table 4 shows that the eigenvalues from 1.01 to 17.71, corresponding to the extraction sums of squared loadings of 13 subscales of the questionnaire. The highest eigenvalue accounted for 17.71% of the total variance, while the lowest eigenvalue explained 1.01%. The result also shows that the extraction eigenvalue does not vary much from the rotation eigenvalue. Therefore, 65 items loaded strongly on 13 factors at the extraction level.

Table S1 in Multimedia Appendix 4 provides the scale construct and its item. All 13 components had acceptable internal reliability (Cronbach α>0.7). All construct eigenvalues were >1. The deleted items were 14 and 15 of domain A, relative advantage, and 56 and 57 of domain D, chief information officer innovativeness, and the deletions were done because their factor loadings were <0.6. After discussing with the experts, questions for items 56 and 57 of domain D, chief information officer innovativeness, were rephrased for clarity. Question for item 15 of domain A, relative advantage, was omitted, and question for item 14 of the same domain remained as it was.

Table S1 in Multimedia Appendix 4 provides a detailed breakdown of various items and their corresponding domain scores, reflecting their relative contribution to each domain. The items are identified by numbers and grouped under distinct domains. For instance, items 18, 19, 22, 20, 21, 16, 17, and 14 display scores ranging from 0.493 to 0.854, indicating their relevance within a particular domain. Similarly, items 63, 65, 62, 64, 61, 60, 59, and 58 exhibit scores from 0.636 to 0.807 and, therefore, were grouped in another domain. The third domain was categorized based on items 40, 42, 44, 43, 41, 38, and 39, with scores ranging from 0.701 to 0.927. The fourth domain includes items 48, 46, 49, 47, 45, and 50, with scores between 0.0715 and 0.829, and the last domain covers items 27, 31, 30, 29, and 28, with scores ranging from 0.773 to 0.822.

Figure 2. Scree plot revealing the 13-factor components extracted from principal component analysis with varimax rotation.

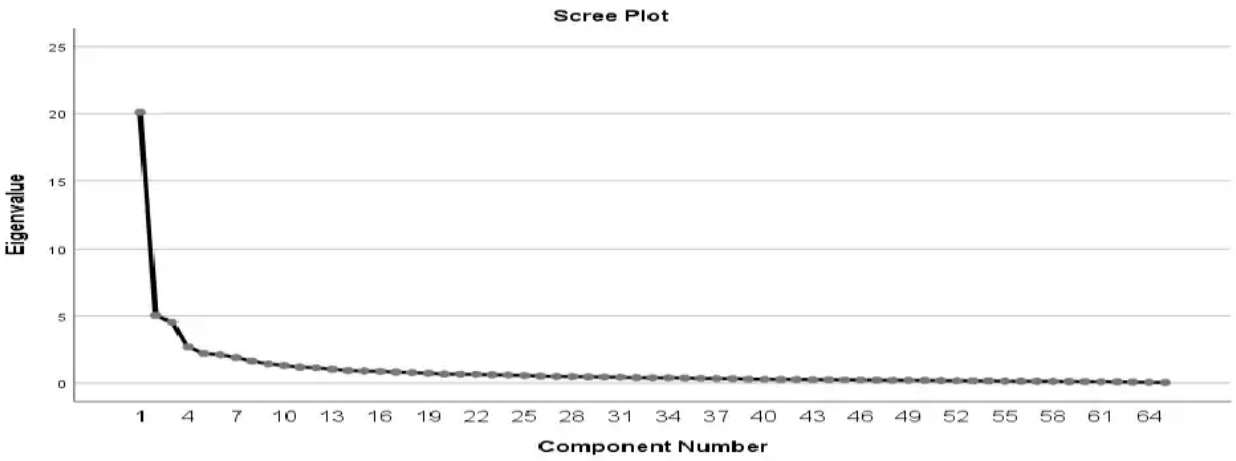


Table 4. The total variance explained.

Factor	Initial eigenvalues			Extraction sums of squared loadings		
	Total, n	Variance (%)	Cumulative (%)	Total, n	Variance (%)	Cumulative (%)
1	18	27.2	27.2	18	27.2	27.2
2	6	8.8	36.1	6	8.8	36.1
3	5	7.6	43.6	5	7.6	43.6
4	5	7.1	50.7	5	7.1	50.7
5	3	5.2	55.9	3	5.2	55.9
6	3	4.3	60.3	3	4.3	60.3
7	2	2.9	63.2	2	2.9	63.2
8	2	2.7	65.9	2	2.7	65.9
9	2	2.6	68.6	2	2.6	68.6
10	1	2.2	70.7	1	2.2	70.7
11	1	2.1	72.8	1	2.1	72.8
12	1	1.7	74.5	1	1.7	74.5
13	1	1.6	76.1	1	1.6	76.1

Discussion

Principal Findings

This study aimed to validate a specific tool used in health care delivery in Malaysia's health care system, a new technology telehealth application called TPC-OHCIS, making this evaluation timely. The 65-item TPC-OHCIS questionnaire demonstrated a comprehensive and reliable tool ready to be used for the assessment of the implementation of the digital application system for monitoring the health service program at the PHCs. Monitoring any health care service performance requires a periodic and ongoing operation to ensure that it can be delivered as planned and with good tracking of progress. Evaluation of how the program should be implemented needs to have a valid tool that can give relevant and appropriate results. A validated tool can be used to replicate similar studies in the future for evaluation and comparison.

Developing and validating a locally customized questionnaire for specific program service delivery is important. Therefore, this study used 4 theories, including the HOT-fit model, DOI theory, TOE theory, and TAM theory. The TPC-OHCIS questionnaire was proven valid based on content, face, and construct validities, as well as good psychometric analysis reliability. A total of 4 scales created using PCA did not align with a local study [17] that assessed the hospital information system implementation in Malaysian hospitals. Measuring implementation of technology in different service set-ups and levels of care, at either hospitals or PHCs, should use a customized tool. No similar research paper is available to check for disagreement or benchmarks on other tools, to compare them with the TPC-OHCIS questionnaire validation. Our findings showed that the 13 scales were validated based on the local population working in the PHC with the TPC-OHCIS. In comparison, an earlier study conducted local testing on the hospital information system and identified 4 scales [18]. Our

analysis focuses on the MCH services, which are the routine and regularly monitoring health services delivered at PHCs.

A CVI score of 0.90 is generally considered very high and indicative of strong content validity for a measurement tool, particularly in the academic context. Content validity refers to the extent to which the items in a measurement instrument represent the entire range of content that the instrument is intended to measure [39]. A CVI of 0.90 provides strong evidence that the measurement tool has undergone rigorous development and validation processes.

Face validity has been widely used not only in questionnaire development but also in questionnaires that were adopted from other researchers' tools, which required local forward-backward translation [37-41]. Experts in the field who were recruited to determine face and content validity in our study gave feedback regarding content relevance, wording, quantity of items, the amount of information, and other related issues. The analysis results for the CVI from 10 experts indicated good quality (>0.9), which complies with the recommendations in the literature [41].

Internal reliability assessment using Cronbach α revealed how close the selected items are to one another when measuring the construct [39]. The Cronbach α value of the 65-item TPC-OHCIS questionnaire was 0.90, which exceeded the suggested minimum value of 0.70 [39]. By measuring Cronbach α, the tool's accuracy and reliability have been verified. Our findings demonstrated that the variables are correlated with the components as constructed through PCA, and it proved that they are internally consistent.

The EFA also determined the TVE for the construct. The TVE illustrated the precision with which the measuring objects and their constituent parts estimate the construct. This study found that the TVE was 76.08%, which exceeded the required minimum of 60%, indicating that the overall variance explained was satisfactory [39]. PCA is the most common EFA method for dimension reduction. The EFA results using varimax rotation with maximum likelihood showed that the TPC-OHCIS questionnaire has 13 subscales.

The study may provide evidence-based recommendations for policy makers and health care stakeholders regarding the implementation and optimization of the TPC-OHCIS in the MCH service delivery monitoring. By identifying key factors influencing the TPC-OHCIS implementation and effectiveness, the study offers actionable insights into policy development, resource allocation, and quality improvement efforts aimed at strengthening MCH care systems. The study may contribute methodological innovations to the field of TPC-OHCIS evaluation, such as novel approaches to questionnaire development, validation, and implementation. By documenting the methodological processes and challenges encountered in assessing the TPC-OHCIS questionnaire validation, the study adds to the methodological toolkit available to researchers and practitioners working in the field of digital health.

Overall, the TPC-OHCIS study offers novel insights into the evaluation of the TPC-OHCIS, which focuses on the MCH services, highlighting the importance of tailored assessment tools, cross-cultural adaptation, policy considerations, and methodological innovations. These insights contribute to a deeper understanding of the complexities and opportunities associated with the integration of digital health technologies into health care delivery and provide a foundation for future research and practice in this area. The TPC-OHCIS questionnaire is validated for evaluating other health information systems, as it takes into account the outcome aspects of health services, technology, organizations, vendors, and human resources. Future researchers can replicate studies to validate the findings of the TPC-OHCIS questionnaire in different populations, settings, and contexts. By replicating studies using a valid and reliable tool with diverse samples, researchers can assess the generalizability and robustness of the questionnaire across various demographic, cultural, and organizational contexts.

Suggestions to Stakeholders

There is a need to integrate machine learning and artificial intelligence into the TPC-OHCIS so that data can be extracted faster and can be used in line with precision medicine. The validated TPC-OHCIS questionnaire can help stakeholders make effective evidence-based decisions in managing patient-centered care. The system needs to be more stable, easier to use, and faster in terms of data collection, without the need for complicated training, and easy to download from the system itself. Awareness related to the importance of complete and accurate data from an organizational perspective needs to be understood at the level of clinic users and data administrators because they will then enter, process, and ensure that the data entered are complete and accurate. These crucial issues were addressed in our study during the content and face validity assessment of the tool.

The COVID-19 pandemic made us see the need for the TPC-OHCIS to be rolled out to wider PHCs in Malaysia. With <10% of the TPC-OHCIS currently implemented, the vision to having a digital health system in Malaysia by the year 2030 needs to be facilitated. The TPC-OHCIS should be developed as a comprehensive package that includes embedded training and health education to facilitate its delivery. The issue of data security can be compared to the concept of banking system apps in Malaysia, where users only need to download the app on their mobile phones to share information with other health care providers, unlike the TPC-OHCIS, which ensures continuity of care. Otherwise, printing a duplicate report is sufficient for the time being while the country works on transitioning all clinics into a digital system.

Study Strength

The study demonstrates the development of a tailored questionnaire specifically designed to assess the implementation process and the effectiveness of the TPC-OHCIS in the context of MCH health services at the primary care level. This tailored approach acknowledges the unique challenges and complexities of MCH service delivery and highlights the importance of designing measurement tools that are sensitive to the needs of this population. The study determined comprehensive items to measure the integration implementation of the TPC-OHCIS, which is one of the digital health technologies focusing on

priority services, such as MCH service delivery at the primary care level. The study sheds light on the practical considerations, barriers, and facilitators associated with the adoption and use of digital health solutions in MCH care settings. The study offers insights into the cross-cultural adaptability assessment tools, particularly in diverse multiethnicity population settings. By evaluating the linguistic and cultural appropriateness of the TPC-OHCIS questionnaire, the study highlights the importance of considering cultural context and language preferences in developing and validating measurement instruments for MCH services.

Limitations of This Study

This study was conducted in 2021 when movement control orders were implemented in Malaysia. The study was conducted entirely online, and all surveys were distributed to the respondents' superiors without proper briefing about the survey and the overall idea of the research except for study information highlighted at the beginning of the online survey questionnaire form. Therefore, each respondent answered as they saw fit, without recognizing the true meaning of the survey. In addition, the deployment of staff during the COVID-19 pandemic could have caused fatigue among them, which was likely to cause them to answer the questions lightly without focus. Meanwhile, the distribution of questionnaires through Google Forms contributed to the risk of bias among respondents when there was no supervision or explanation that could be given as a guide to answering the questionnaire.

Conclusions

The psychometric validation process of the questionnaire was done comprehensively for all iterative stages, including initial reliability tests, potential modifications based on these tests, and integration with existing digital health assessment frameworks. The novelty of the study lies in its unique approach to conceptualizing the implementation of the TPC-OHCIS. This may offer new insights into the integration of life stage records, with a focus on priority services at PHCs, such as MCH services, for its development. Cross-cultural adaptability was considered to ensure wider applicability, and the rigor of the process was demonstrated in part by high CVI scores, which is important in an academic research setting. Likert scale points were carefully chosen to capture nuanced responses, and while the TPC-OHCIS has certain academic limitations, its robustness paves the way for future research on policy implications and telehealth. Comparison with existing literature establishes its validity and reliability, showing its potential impact on policy development.

Acknowledgments

The authors would like to especially thank the Selangor Health State Department, Kuala Lumpur and Putrajaya Health State Department, and Pahang Health State Department for their support in the study. The authors would also like to thank the Director General of Health Malaysia for his permission to conduct the study and publish this paper. This work was supported by a research grant from the Ministry of Higher Education (FRGS/1/2021/SKK04/UKM/02/1) and by the Medical Faculty of Universiti Kebangsaan Malaysia, which contributed to the publication costs of this paper.

Authors' Contributions

SI, RS, and RI conceived the scope of the review to meet the study objectives. Literature search, data collection, analysis, and interpretation were conducted by SI and verified by RS. SI drafted the first iteration of the manuscript, which was revised and improvised by RS. All authors made substantial contributions to the critical review, editing, and revision of the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Evidence table.
[PDF File (Adobe PDF File), 151 KB - humanfactors_v12i1e53630_app1.pdf]

Multimedia Appendix 2
Tele-Primary Care Oral Health Clinical Information System questionnaire.
[DOCX File , 42 KB - humanfactors_v12i1e53630_app2.docx]

Multimedia Appendix 3
Expert review analysis.
[XLSX File (Microsoft Excel File), 22 KB - humanfactors_v12i1e53630_app3.xlsx]

Multimedia Appendix 4
Domain construct according to item question.
[DOCX File , 25 KB - humanfactors_v12i1e53630_app4.docx]

References

1. Yang K, Hu Y, Qi H. Digital health literacy: bibliometric analysis. J Med Internet Res 2022 Jul 06;24(7):e35816 [FREE Full text] [doi: 10.2196/35816] [Medline: 35793141]
2. Marcolino MS, Diniz CS, Chagas BA, Mendes MS, Prates R, Pagano A, et al. Synchronous teleconsultation and monitoring service targeting COVID-19: leveraging insights for postpandemic health care. JMIR Med Inform 2022 Dec 22;10(12):e37591 [FREE Full text] [doi: 10.2196/37591] [Medline: 36191175]
3. Abd Ghani MK, Mohamed MA, Mostafa SA, Mustapha A, Aman H, Jaber MM. The design of flexible telemedicine framework for healthcare big data. Int J Eng Technol 2018;7:461-468. [doi: 10.14419/ijet.v7i3.30.19096]
4. Abidi SS, Goh A, Yusoff Z. Telemedicine and medical informatics in the multimedia super corridor: the Malaysian vision. Stud Health Technol Inform 1998;52 Pt 2:1282-1286. [Medline: 10384666]
5. Mudaris IS. Digitalised health records: does Malaysia need it? Khazanah Research Institute. 2023 Mar 27. URL: https://www.krinstitute.org/Views-@-Digitalised_Health_Records-;_Does_Malaysia_Need_It%5E.aspx [accessed 2024-05-24]
6. Cresswell K, Williams R, Sheikh A. Developing and applying a formative evaluation framework for health information technology implementations: qualitative investigation. J Med Internet Res 2020 Jun 10;22(6):e15068 [FREE Full text] [doi: 10.2196/15068] [Medline: 32519968]
7. Bhattacharya AA, Umar N, Audu A, Felix H, Allen E, Schellenberg JR, et al. Quality of routine facility data for monitoring priority maternal and newborn indicators in DHIS2: a case study from Gombe State, Nigeria. PLoS One 2019 Jan 25;14(1):e0211265 [FREE Full text] [doi: 10.1371/journal.pone.0211265] [Medline: 30682130]
8. Al-Shorbaji N. Improving healthcare access through digital health: the use of information and communication technologies. In: Agrawal A, Kosgi S, editors. Healthcare Access. London, UK: IntechOpen; 2022.
9. Ibrahim AA, Ahmad Zamzuri M, Ismail R, Ariffin AH, Ismail A, Muhamad Hasani MH, et al. The role of electronic medical records in improving health care quality: a quasi-experimental study. Medicine (Baltimore) 2022 Jul 29;101(30):e29627 [FREE Full text] [doi: 10.1097/MD.0000000000029627] [Medline: 35905245]
10. Fryer K, Delgado A, Foti T, Reid CN, Marshall J. Implementation of obstetric telehealth during COVID-19 and beyond. Matern Child Health J 2020 Sep;24(9):1104-1110 [FREE Full text] [doi: 10.1007/s10995-020-02967-7] [Medline: 32564248]
11. Yuen J, Pike S, Khachikyan S, Nallasamy S. Telehealth in ophthalmology. In: Linwood SL, editor. Digital Health. Brisbane, Australia: Exon Publications; 2022.
12. Samiei V, Wan Puteh SE, Abdul Manaf MR, Abdul Latip K, Ismail A. Are Malaysian diabetic patients ready to use the new generation of health care service delivery? A telehealth interest assessment. Malays J Med Sci 2016 Mar;23(2):44-52 [FREE Full text] [Medline: 27547114]
13. Akiyama H, Hasegawa Y. A trial case of medical treatment for primary headache using telemedicine. Medicine (Baltimore) 2018 Mar;97(9):e9891 [FREE Full text] [doi: 10.1097/MD.0000000000009891] [Medline: 29489688]
14. Dehnavieh R, Haghdoost A, Khosravi A, Hoseinabadi F, Rahimi H, Poursheikhali A, et al. The District Health Information System (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf Manag 2019 May 13;48(2):62-75. [doi: 10.1177/1833358318777713] [Medline: 29898604]
15. Faujdar DS, Sahay S, Singh T, Jindal H, Kumar R. Public health information systems for primary health care in India: a situational analysis study. J Family Med Prim Care 2019 Nov;8(11):3640-3646. [doi: 10.4103/jfmpc.jfmpc_808_19] [Medline: 31803666]
16. Beheshti L, Kalankesh LR, Doshmangir L, Farahbakhsh M. Telehealth in primary health care: a scoping review of the literature. Perspect Health Inf Manag 2022 Jan 1;19(1):1n [FREE Full text] [Medline: 35440933]
17. World Health Organization. Regional Office for the Western Pacific. Malaysia health system review. World Health Organization Regional Office for the Western Pacific. 2012. URL: https://iris.who.int/handle/10665/206911 [accessed 2024-05-24]
18. Marzuki NM, Ismail S, Abdul Mohsein NA, Ehsan FZ. Evaluation of telehealth implementation in government primary health clinics - a study protocol. In: Proceedings of the International Conference on Green and Ubiquitous Technology. 2012 Presented at: GUT 2012; July 7-8, 2012; Bandung, Indonesia. [doi: 10.1109/gut.2012.6344169]
19. Malaysia's telemedicine blueprint: leading healthcare into the information age. Ministry of Health Malaysia. 1997 Jul 25. URL: https://www.moh.gov.my/moh/resources/auto%20download%20images/5ca1b20928065.pdf [accessed 2024-04-24]
20. Mohan J, Razali Raja Yaacob R. The Malaysian Telehealth Flagship Application: a national approach to health data protection and utilisation and consumer rights. Int J Med Inform 2004 Mar 31;73(3):217-227. [doi: 10.1016/j.ijmedinf.2003.11.023] [Medline: 15066550]
21. Maon SN, Edirippulige S. An overview of the national telehealth initiative in Malaysia. Stud Health Technol Inform 2010;161:95-103. [Medline: 21191162]
22. Teleprimary Care – Oral Health Clinical Information System (TPC-OHCIS). Bahagian Pembangunan Kesihatan Keluarga. URL: https://hq.moh.gov.my/bpkk/index.php/sistem-tpc-ohcis/sistem-tpc-ohcis [accessed 2024-04-24]
23. Teleprimary care and oral health clinical information system. MIMOS. 2020. URL: https://www.mimos.my/wp-content/uploads/2020/09/TPC-OHCIS.pdf [accessed 2024-04-24]
24. Teleprimary Care and Oral Health Clinical Information System (TPC-OHCIS). MIMOS. URL: https://www.mimos.my/main/teleprimary-care-and-oral-health-clinical-information-system-TPC-OHCIS [accessed 2024-04-24]

25. Marzuki N, Ismail S, Al-Sadat N, Ehsan FZ, Chan CK, Ng CW. Integrating information and communication technology for health information system strengthening: a policy analysis. Asia Pac J Public Health 2015 Nov;27(8 Suppl):86S-93S. [doi: 10.1177/1010539515590180] [Medline: 26085477]
26. Aziz A, Zork N, Aubey JJ, Baptiste CD, D'Alton ME, Emeruwa UN, et al. Telehealth for high-risk pregnancies in the setting of the COVID-19 pandemic. Am J Perinatol 2020 Jun;37(8):800-808 [FREE Full text] [doi: 10.1055/s-0040-1712121] [Medline: 32396948]
27. Wu H, Sun W, Huang X, Yu S, Wang H, Bi X, et al. Online antenatal care during the COVID-19 pandemic: opportunities and challenges. J Med Internet Res 2020 Jul 22;22(7):e19916 [FREE Full text] [doi: 10.2196/19916] [Medline: 32658860]
28. Parajuli R, Bohara D, Kc M, Shanmuganathan S, Mistry SK, Yadav UN. Challenges and opportunities for implementing digital health interventions in Nepal: a rapid review. Front Digit Health 2022;4:861019 [FREE Full text] [doi: 10.3389/fdgth.2022.861019] [Medline: 36120714]
29. Faizal AM, Muhammad Azrai A, Abdul Kadir AK. The impact of COVID-19 pandemic on oncofertility services in advanced reproductive centre UKM Medical Centre Malaysia. Asian Pac J Cancer Care 2020 Jul 01;5(S1):165-166. [doi: 10.31557/apjcc.2020.5.s1.165-166]
30. Lindberg MH, Venkateswaran M, Abu Khader K, Awwad T, Ghanem B, Hijaz T, et al. eRegTime, efficiency of health information management using an electronic registry for maternal and child health: protocol for a time-motion study in a cluster randomized trial. JMIR Res Protoc 2019 Aug 07;8(8):e13653 [FREE Full text] [doi: 10.2196/13653] [Medline: 31392962]
31. Khobi JA, Mtebe JS, Mbelwa JT. Factors influencing district health information system usage in Sierra Leone: a study using the technology-organization-environment framework. Electron J Inf Syst Dev Ctries 2020 Jun 08;86(6):e12140. [doi: 10.1002/isd2.12140]
32. Alsharo M, Alnsour Y, Alabdallah M. How habit affects continuous use: evidence from Jordan's national health information system. Inform Health Soc Care 2020 Jan;45(1):43-56. [doi: 10.1080/17538157.2018.1540423] [Medline: 30457025]
33. Salleh MI, Abdullah R, Zakaria N. Evaluating the effects of electronic health records system adoption on the performance of Malaysian health care providers. BMC Med Inform Decis Mak 2021 Feb 25;21(1):75 [FREE Full text] [doi: 10.1186/s12911-021-01447-4] [Medline: 33632216]
34. Ahmadi H, Nilashi M, Ibrahim O. Organizational decision to adopt hospital information system: an empirical investigation in the case of Malaysian public hospitals. Int J Med Inform 2015 Mar;84(3):166-188. [doi: 10.1016/j.ijmedinf.2014.12.004] [Medline: 25612792]
35. Ahmadi H, Nilashi M, Shahmoradi L, Ibrahim O, Sadoughi F, Alizadeh M, et al. The moderating effect of hospital size on inter and intra-organizational factors of Hospital Information System adoption. Technol Forecast Soc Change 2018 Sep;134:124-149. [doi: 10.1016/j.techfore.2018.05.021]
36. Zhang X, Yu P, Yan J, Ton A M Spil I. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15(1):71 [FREE Full text] [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
37. Taherdoost H. What is the best response scale for survey and questionnaire design; review of different lengths of Rating Scale / Attitude Scale / Likert Scale. Int J Acad Res Manag 2019;8(1):1-10 [FREE Full text] [doi: 10.4135/9781412963947.n273]
38. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
39. Taherdoost H. Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. SSRN 2016 Aug 10 [FREE Full text] [doi: 10.2139/ssrn.3205040]
40. Suhr D. Exploratory or confirmatory factor analysis? In: Proceedings of the 31st Annual SAS Users Group International Conference. 2006 Presented at: SUGI 2006; March 26-29, 2006; San Francisco, CA.
41. Yusoff MS. ABC of response process validation and face validity index calculation. Educ Med J 2019 Oct 31;11(3):55-61. [doi: 10.21315/eimj2019.11.3.6]
42. Yusoff MS. Construct validity, internal consistency and normative data of the USMaP-i in a sample of medical students. Int Med J 2013 Feb;20(1):1-7 [FREE Full text]
43. Svedbo Engström M, Leksell J, Johansson UB, Eeg-Olofsson K, Borg S, Palaszewski B, et al. A disease-specific questionnaire for measuring patient-reported outcomes and experiences in the Swedish National Diabetes Register: development and evaluation of content validity, face validity, and test-retest reliability. Patient Educ Couns 2018 Jan;101(1):139-146. [doi: 10.1016/j.pec.2017.07.016] [Medline: 28736071]

Abbreviations

CVI: content validity index
DOI: diffusion of innovation
EFA: exploratory factor analysis

HCW: health care worker
HOT-fit: human organization technology-fit
MCH: maternal and child health
PCA: principal component analysis
PHC: primary health care clinic
S-CVI: scale-level content validity index
TAM: technology acceptance model
TOE: technology-organization-environment
TPC-OHCIS: Tele-Primary Care Oral Health Clinical Information System
TVE: total variance explained

Edited by A Choudhury; submitted 12.10.23; peer-reviewed by F Medina, Y Chu, D Chrimes; comments to author 15.03.24; revised version received 14.05.24; accepted 25.11.24; published 29.01.25.

Please cite as:
Sutan R, Ismail S, Ibrahim R
Evaluating the Development, Reliability, and Validation of the Tele-Primary Care Oral Health Clinical Information System Questionnaire: Cross-Sectional Questionnaire Study
JMIR Hum Factors 2025;12:e53630
URL: https://humanfactors.jmir.org/2025/1/e53630
doi:10.2196/53630
PMID:

Original Paper

Creation of Text Vignettes Based on Patient-Reported Data to Facilitate a Better Understanding of the Patient Perspective: Design Study

Sofia Charlotte Kelly[1*], MSc; Emilie Kauffeldt Wegener[1*], MA, MSc; Lars Kayser[1*], MD, PhD

Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[*]all authors contributed equally

Corresponding Author:
Sofia Charlotte Kelly, MSc
Department of Public Health
University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1353
Denmark
Phone: 45 61722415
Email: sofiakelly95@gmail.com

Abstract

Background: Patient-reported outcome (PRO) data refer to information systematically reported by patients, or on behalf of patients, without the influence of health care professionals. It is a focal point of the health care system's ambition toward becoming more involving and personalized. It is recognized that PROs provide valuable data. However, despite this recognition, there are challenges related to both patients' and clinicians' accurate interpretations of the quantitative data. To overcome these challenges, this study explores text vignettes as a representation of PROs.

Objective: This study aimed to develop data-informed text vignettes based on data from the Readiness and Enablement Index for Health Technology (READHY) instrument as another way of representing PRO data and to examine how these are perceived as understandable and relevant for both patients and clinicians.

Methods: The text vignettes were created from participant responses to the READHY instrument, which encompasses health literacy, health education, and eHealth literacy. The text vignettes were created from 13 individual text strings, each corresponding to a scale in the READHY instrument. This study consisted of 3 sequential parts. In part 1, individuals with chronic obstructive pulmonary disease completed the READHY instrument, providing data to be used to create vignettes based on cluster profiles from the READHY instrument. Part 2 focused on the development of scale-based strings representing all READHY dimensions, which were evaluated through iterative cognitive interviews. In part 3, clinicians and patients assessed the understanding and relevance of the text vignettes.

Results: Clinicians and patients both understood and related to the text vignettes. Patients viewed the text vignettes as an accurate reflection of their PRO responses, and clinicians perceived the text vignettes as aligned with their understanding of patients' experiences.

Conclusions: Text vignettes can be developed using PRO instruments, with individual scales as input strings. This provides an opportunity to present numeric values in a text format that is understandable and recognizable to most patients and clinicians. Challenges with the vignette's language and layout require customization and clinician training to ensure meaningful interpretation. Findings also support the need to expand the study and enhance clinical relevance with alternative or contextually relevant text vignettes.

(JMIR Hum Factors 2025;12:e58077) doi:10.2196/58077

KEYWORDS

patient-reported outcome; text vignette; data representation; Readiness and Enablement Index for Health Technology; understanding; health care system; data analysis; clinical training; clinician; health professional

Introduction

Background

Patient-reported outcome (PRO) data represent patients' self-assessed health status in terms of physical and mental health and do not involve health professionals [1]. PRO data can be used as a screening tool before outpatient visits as a supplement to in-person or virtual consultations, serving as a stratification tool to reduce the number of visits needed. PRO data can also be used during consultations to screen for symptoms or serve as a dialogue and decision tool. The use of PRO data during consultations may increase the patient's self-awareness and health literacy, facilitate joint decision-making, enable self-management, and lead to better health outcomes [1,2].

Health Literacy

Competence in health literacy and digital knowledge is fundamental for an accurate understanding of PRO data as well as knowledge about health behavior and health practices and how patients understand and interpret health information [3]. Challenges arise when data accessibility does not consider differences in health literacy. While aggregate-level understanding may tolerate some variations, individual patient pathways require a common understanding to capture nuances [4,5]. Therefore, it is crucial to account for patients' health literacy and cognition in the clinical use of PROs.

Interpretation of PRO Data

When incorporating PROs into individual patient pathways, it becomes crucial to ensure that patients fully understand the questions, fostering a shared understanding. In working with PRO data, an assumption is often made that people have the same needs, preferences, perceptions, and experiences and will attach similar importance to items in the questionnaire, thus leaving no room to capture different additions of value or uncover nuances [4]. The quantitative part of PRO data operates within a set of assumptions about the world, which states that the standardized numerical score is the objective representation of the respondents' health status and health-related quality of life, independent of an individual's unique life story [4].

For both clinicians and patients, it may be difficult to understand the meaning of PRO data when they are presented as values. Patients often prefer visual representations, such as bar charts and line charts, which prove useful for quick comprehension and comparisons over time [6]. However, the accuracy of interpretation is closely tied to the individuals' prior knowledge of these graphical representations [6].

Graphical presentations may be a challenge. Grossman et al [7] highlighted that individuals lacking prior knowledge of bar charts struggled to interpret them accurately. Furthermore, health literacy plays a crucial role in correct interpretation, with studies suggesting that patients with lower health literacy may find it challenging to grasp the longitudinal nature of PRO data [8]. There is a notable correlation between educational level and accurate interpretation, along with age-related difficulties in correctly interpreting graphic representations of PRO data [6,9]. Patients generally, regardless of educational level, prefer less-complex presentations of PRO data, but a correlation has been demonstrated between age and accurate interpretation, where patients aged >65 years had substantially more difficulty interpreting graphic representations of PRO data correctly [6,9]. In Denmark, 1 in 5 individuals is aged >65 years [10], which means that a relatively large proportion of all individuals will potentially have difficulty interpreting PRO data.

The use of traffic light colors, that is, green, yellow, and red, in representing PRO data adds another layer of complexity. While this method increases accuracy in interpretation across various graphical representations [6,11], it is not foolproof and 93% of patients in a study population were able to correctly interpret the meaning of traffic light colors [6]. Despite this visual aid, there remains a need for additional descriptive explanations to enhance understanding [9].

Patients' and health care professionals' perceived understanding of graphically presented PRO data often exceeds their actual comprehension [9]. In addition, studies indicate that there may be a discrepancy between patients' preferred form of visualization and the form of visualization they can interpret with the highest accuracy [9].

Patients may also view the use of PRO data in treatment as impersonal and nonbeneficial for the relationship between themselves and their clinicians because they experience that the focus shifts from their own perspective to quantitative goals [12]. This perspective highlights the significance of maintaining a patient-centered approach.

A Common Understanding of PROs

When patients and clinicians collaborate on data interpretation, potential errors may arise due to questionnaire misinterpretation. This divergence in understanding creates a need for a common language or understanding of the representations of colors, figures, or text. A challenge is that patients and clinicians possess distinct cognitive frameworks influencing their actions and comprehension [13]. Data interpretation involves understanding available data, extracting information, and supplementing it for comprehensive comprehension [13]. Clinicians are shaped by a positivist background and may differ from patients in cognitive approaches, impacting data interpretation. When PRO data are translated into quantitative targets, it may be unclear what data have been made available and what information can be extracted. Patients provide a wide range of data, but data visualization often only includes part of these data.

Despite psychometric validation, there is a need to address discordance in individual interpretations of PRO data. Discordance reasons include word interpretation mismatches, evolving patient circumstances, and differing expectations [14]. Awareness of these differences is crucial in interpreting PRO data [14]. PRO tools vary in complexity, leading to further potential errors if patients lack an accurate understanding. A study found that only 11% of 59 PRO questionnaires are readable by the average UK adult [15], emphasizing the need for a patient-clinician dialogue to enhance understanding and decision-making [5]. Numerical scores derived from PROs can represent a person's health status or health-related quality of life to a limited extent because some of the narrative is lost

along the way. In the representation of PRO data, there is a need to be able to give meaning to individual experiences [4]. Positivism argues that knowledge should be based on the objective world via empirically verified causal explanations based on "positive" facts and not on abstract inferences; however, according to Meadow [4], we cannot understand the patient if we do not also try to understand subjective and emotional meaning.

PROs as Text Vignettes

One approach to a more understandable and personalized representation of data about a group of patients or individuals is vignettes created based on a combination of clustered data, which are enriched with sociodemographic characteristics and interviews with individuals representing these clusters. An example of this is the Optimizing Health Literacy and Access (OPHELIA) process, where issues related to health care access and engagement are identified and addressed [16,17]. This way of creating data-informed text vignettes is a powerful tool to provide clinicians and policy makers with insights into a more detailed description of archetypes of patients [18]. In this way, text vignettes play a pivotal role in providing a nuanced understanding of the perceived status of various subgroups of patients in a specific context or area. The primary aim is to bring to life the "person behind the numbers," referring to PRO scores, addressing the challenge of visualizing the individuals behind the statistical data. To the best of our knowledge, this has not yet been done at the level of individuals by creating person-specific text vignettes.

In this study, inspired by the OPHELIA process, we will explore whether text vignettes can be developed from more structured text inputs and how this work can be advanced to create text vignettes that can be used at the individual level, rather than at the group level, to present personalized texts that reflect each patient's perception of their condition. The study uses the Readiness and Enablement Index for Health Technology (READHY) [19], which covers 3 key themes—self-management, social capital, and digital health literacy. This instrument was selected because, as its developers, we have extensive experience with its performance, and its design makes it well suited as a PRO tool for telehealth services. The study demonstrates the process of converting psychometric instrument scales into data-informed text vignettes, highlighting the challenges and implications to consider for ensuring understanding and relevance for both patients and clinicians.

Objectives

This study aimed to develop data-informed text vignettes based on data from the READHY instrument as another way of representing PRO data and to examine how these are perceived as understandable and relevant for both patients and clinicians.

Part 1 Introduction: Generating Vignettes Based on READHY Clusters Enriched With Sociodemographic Data

This part is a proof of concept, where cluster-based profiles were created to help develop representatives of characteristics of a patient population divided into subgroups. This enabled us to work with materials based on real-life data varying with respect to the 13 scale scores. This has provided insights into how dimension names can be phrased for text vignettes without a systematic approach, while also helping us establish a shared understanding of the OPHELIA [16] principles. These principles were operationalized in a data-informed manner, bridging the previously reported qualitative approach with a more quantitative-based methodology.

Part 2 Introduction: Creation of Text Strings and Testing of Meaning in Vignettes

In part 2, scale-based text strings describing the different dimensions measured in READHY were developed through a series of iterations. These were then combined into comprehensive text vignettes profiling the participants. For this process, the upper threshold was adjusted from 2.50 to 2.70 based on data from the study by Kayser et al [20]. The vignettes were continuously refined through cognitive interviews to evaluate and improve the participants' understanding and ability to comprehend the content.

Part 3 Introduction: Applicability of Text Vignettes in a Clinical Setting

The text strings iteratively created and validated in part 2 served as the foundation for the work in part 3. In part 3, the focus shifted to testing the applicability of individual text vignettes within a clinical setting. The emphasis was on how meaningful and relevant these text vignettes were in a setting involving both patients and clinicians within a telehealth service, where the READHY instrument was suitable as a PRO instrument. The clinical setting included here was the same as the one used to gather data for part 1.

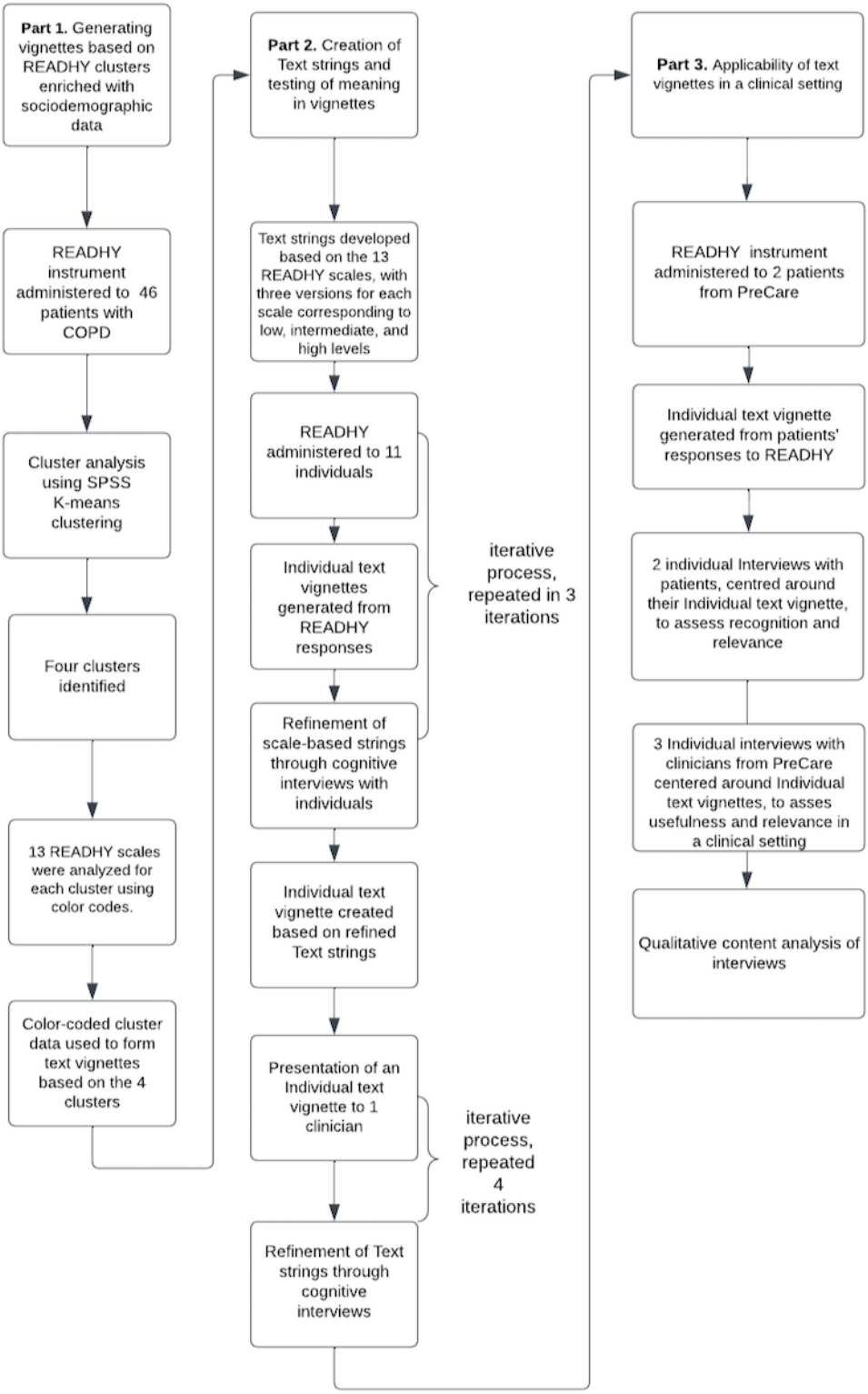
Methods

Overview

The study comprises 3 parts, each building upon the previous one. In part 1, which was inspired by the OPHELIA process [16], the READHY instrument was administered to individuals living with chronic obstructive pulmonary disease (COPD). Data gathered from this group of participants served as the foundation for generating 4 clusters, which were characterized by high, intermediate, and low levels of the 13 READHY scales, which together with another Health Literacy Questionnaire (HLQ) scale, "actively managing my health and sociodemographics", were used to develop a proof of concept of how data-informed text vignettes could be used to illustrate the quantitative data qualitatively to better understand profiles of user groups. The text vignettes in part 2 were crafted as narratives, emphasizing the group's high or low scores. This approach created profiles that highlighted their resources and barriers. In part 2, we expanded on the idea by constructing scale-specific strings for each of the 3 levels for each scale in READHY. The text strings were created to constitute a list of variables, which in a later process could be combined based on an algorithm to form the text vignettes. This computer-based combination of the text strings is beyond the scope of this study that only addresses the development of our concept up to a technology readiness level of 3 [21]. Text vignettes are consequently created by manually combining the text strings.

In this process, the text strings were evaluated for clarity and understanding through multiple iterations of cognitive interviews with a diverse set of participants. In part 3, the refined scale-based strings were combined manually, informed by the 2 patients' READHY scores, into individual text vignettes, which were presented to clinicians and patients for assessment of their understanding, relevance, and recognizability (Figure 1).

Figure 1. Flowchart: the 3-part process for examining the use of text vignettes. COPD: chronic obstructive pulmonary disease; READHY: Readiness and Enablement Index for Health Technology.



Setting and Population

The setting for the reported activities was in part 1 and part 3 of a telehealth service, PreCare [22], which provides 24-7 access to a response and coordination center manned with registered nurses and backed up by an e-doctor. PreCare is organized based on the Epital care model [22,23]. The participants in part 2 were

recruited broadly among individuals in the eastern part of Denmark and clinicians not currently working in telehealth.

Participants

Part 1

The participants were 46 adults living with COPD, with a mean age of 74.4 (SD 6.8; range 59-91) years. They were patients recruited by convenience sampling from the European Commission funded project Smart Inclusive Living Environments, were enrolled in the PreCare project in Region Zealand, and were diagnosed with different severity levels of COPD [24]. The inclusion criteria were a COPD diagnosis and being active in the PreCare project. The patients should be able to understand and communicate in Danish.

Part 2

Individuals

A total of 11 individuals were recruited using a snowball technique in the Capital Region of Denmark and Region Zealand. Initially, 3 members of interest were identified and approached within their local areas. These individuals were then asked to refer others in their community who might be willing to participate. With their consent, we received the necessary contact information and subsequently reached out to the referred participants by phone to arrange their involvement in the study. The inclusion criteria for individuals' participation was that they had to be aged >18 years, able to understand Danish, capable of giving informed consent, and able to complete the READHY instrument, with assistance if needed (ie, the ability to read and write independently was not a criterion). The first 3 recruited individuals resided in the metropolitan area, had medium-length higher education (International Standard Classification of Education level 6-7) [25], and had an average age of 53 (SD 12.36) years. This group was expanded to include 8 participants, aged between 36 and 86 years, from the capital region, Lolland, and Jutland with educational backgrounds ranging from International Standard Classification of Education level 4 to 6 [25] to attain a more diverse and comprehensive perspective.

Clinicians

A total of 4 clinicians were also recruited using a snowballing technique. Initially, 3 clinicians were approached, but not all expressed interest in participating. These contacts referred to others, and with their consent, we obtained their contact information and contacted them. The clinicians who participated in the interviews were 1 registered nurse, 2 physicians, and 1 professional with a master's and PhD degree in biomedicine. Of the 4 physicians, 2 (50%) had prior knowledge of READHY as active researchers using the instrument. All physicians were involved in professional settings related to digital health in patient care, and they were all familiar with PROs.

Part 3

A total of 3 registered nurses and 2 patients from the PreCare clinic were interviewed about the understanding, recognizability, and relevance of individual text vignettes in a clinical setting.

The 2 patients, a woman aged 75 years and a man aged 80 years, lived in the region where the PreCare project is based. They were selected by PreCare staff based on the following inclusion criteria: a COPD diagnosis, active participation in PreCare, the ability to understand and communicate in Danish, clinic enrollment at the time of data collection, and prior consent through PreCare as well as informed consent specifically for this study. A total of 3 (75%) of the 4 nurses in PreCare were available for interviews.

Application of the READHY Instrument

READHY is a validated PRO instrument that combines several aspects of PROs to illuminate experienced support, digital health literacy, and coping with everyday life in relation to one's health [19]. The tool focuses on aspects of knowledge and skills related to health; self-care, disease, and health-related mindset; experiences with health technology and the individual understanding of these; as well as the extent to which users feel supported by relatives and health care professionals [19].

READHY has been used to identify different health technology readiness profiles within different patient groups, such as patients living with type 2 diabetes and patients with cancer [26,27].

READHY consists of 65 items comprising 13 dimensions or scales. The 13 scales are selected from 3 PRO instruments. Four scales are from the Health Education Impact Questionnaire (heiQ), which evaluates the effect of health education interventions on self-care [28]. These address aspects of self-management with relevance for a digital health context, such as insights and attitudes. Two scales from the HLQ, which assesses health literacy, including support from relatives, peers, and health professionals [29], and all 7 scales from the eHealth Literacy Questionnaire (eHLQ), which address patients' digital health literacy, were included [30]. The eHLQ informs about knowledge, skills, motivation, perceptions, and experiences in relation to digital health. The 13 dimensions of READHY, which correspond to its 13 scales, are presented in Textbox 1.

Textbox 1. The 13 dimensions of the Readiness and Enablement Index for Health Technology framework.

- Health Education Impact Questionnaire (heiQ) 3: self-monitoring and insight
- heiQ4: constructive attitudes and approaches
- heiQ5: skill and technique acquisition
- heiQ8: emotional distress
- Health Literacy Questionnaire (HLQ) 1: feel understood and supported by health care providers
- HLQ4: social support for health
- eHealth Literacy Questionnaire (eHLQ) 1: using technology to process health information
- eHLQ2: understanding of health concepts and language
- eHLQ3: ability to actively engage with digital services
- eHLQ4: feel safe and in control
- eHLQ5: motivated to engage with digital services
- eHLQ6: access to digital services that work
- eHLQ7: digital services that suit individual needs

The READHY instrument uses a 4-point response scale for all items in the instrument [19], where 1 represents strongly disagree, 2 represents disagree, 3 represents agree, and 4 represents strongly agree. The results are 13 scale scores. For the scale heiQ8 (emotional distress), which includes 6 items, the scale is reversed by subtracting the scored value from 5. This adjustment simplifies interpretation, as higher scores now correspond to less impact.

Ethical Considerations

Written informed consent was obtained from all participants after providing information in accordance with the Helsinki II Declaration, including their right to opt out. Participants did not receive any salary or reimbursement of costs. Health science questionnaire surveys and interview studies that do not involve human biological material (section 14(2) of the Danish Act on Committees) do not require reporting or approval from the Danish National Center for Ethics [31].

Part 1 Methods

Data Sampling

The READHY instrument was handed out and collected with a combination of mail using prestamped envelopes by EKW and a researcher in PreCare in May 2021. A total number of 120 surveys including the READHY instrument were distributed.

Cluster Analysis

In managing the READHY data from the 46 patients, clusters were created using the K-means clustering function in IBM's software package SPSS Statistics (version 24) [32]. Three models were created with 3, 4, and 5 clusters. On the basis of an evaluation of which cluster model would be most meaningful from a clinical perspective, a 4-cluster model was selected. The final cluster constituted 41 valid cases and 5 missing cases out of the 46 responses. The number of cases in each cluster was as follows: 16 in cluster 1, 5 in cluster 2, 11 in cluster 3, and 9 in cluster 4.

The cluster data were then used to form the group-based text vignettes. For each cluster, the 13 scales were initially labeled with a color indicating whether the level was low, intermediate, or high. We used arbitrary thresholds informed by a regional survey, Sundhedsprofilen 2021 [33], which is the region where the PreCare project is based. All values <2.00 were assigned red, all values between 2.01 and 2.50 were assigned yellow, and values >2.50 were given green. This helped us create text strings that were specific for each of the colors indicating high or low average scores for each scale and assigning them to the cluster text vignette. The text strings were created from the dimension name and informed by the items in the scale. Additional data from an additional HLQ scale, "actively managing my health," and sociodemographic variables, such as educational level, experience living with, and number of chronic conditions, were added to the profiles to enrich them and create personas that were recognizable to the clinicians.

Part 2 Methods

Overview

The overall process for constructing the text vignettes involved several steps. First, text strings were developed based on the 13 READHY scales, with 3 versions for each scale corresponding to low, intermediate, and high levels. Next, participants completed the instrument. The average score for each scale was then manually calculated from their responses, and this average score was matched with the corresponding text string for that scale score. Finally, the 13 text strings were combined into an individual text vignette. The text strings were combined in a coherent order, prioritizing readability and clarity, rather than strictly following the numerical order of the scales. In part 2, the focus was only on the READHY scales and text vignettes; therefore, part 2 did not include sociodemographic information, as these are anticipated to be retrieved from other data sources such as electronic health records as they are specific at the individual level.

Constructing Data-Informed Text Vignettes

The text vignettes were created by combining 13 text strings into a cohesive PRO profile. Each sentence in the text vignette corresponds to a specific scale in READHY, describing the knowledge, skills, and attitudes that the scale items aim to capture.

Initially, 39 text strings were constructed—3 for each of the 13 READHY scales, representing different levels of knowledge, skills, or attitudes within each scale. These initial text strings were closely aligned with the scale names, as the scale names summarized the experiences the items were designed to encompass. For example, the scale "self-monitoring and insight" (heiQ3) assesses an individual's ability to monitor their health and their insight into their health [19]. An initial sentence for this scale might read as follows: "Has good insight into and control over his/her health and has a constructive attitude and approach to it."

The levels are determined based on the average score for each scale, with 3 categories: level 1 indicates problematic knowledge, skills, and attitudes; level 2 corresponds to limited or insufficient knowledge, skills, and attitudes; and level 3 signifies adequate knowledge, skills, and attitudes. To be classified as level 3, the average score must be >2.7, indicating agreement or strong agreement with the statement (Table 1). The cutoffs are estimates defined by the authors and have not been validated by other parties. However, they have been used and found relevant in other studies using HLQ and eHLQ scores [20,33].

Table 1. Breakdown of knowledge, skills, and attitude into levels based on average scores.

Average scale score[a]	Level of knowledge, skills, and attitudes
<2.00	Level 1
2 to 2.7	Level 2
>2.7	Level 3

[a]The cutoffs are author-defined estimates, unvalidated by others but found relevant in studies using the Health Literacy Questionnaire and eHealth Literacy Questionnaire [20,33].

These 39 text strings are organized in Table 2 with the 13 scales as rows and the 3 levels as columns. To construct a text vignette, the appropriate text strings were selected from the table based on the participant's average scale score for each scale. The 13 selected text strings were then manually combined into a coherent text vignette. The initial text strings were revised throughout part 2 of the study in an iterative process using cognitive interviews for accuracy and relevance.

Table 2. Refined text strings based on scores.

Scale	Scores <2.00	Scores between 2.00 and 2.70	Scores >2.70
heiQ3[a]: self-monitoring and insight	He/she experiences having insufficient insight into his/her state of health, with insufficient knowledge of how to deal with health problems when they arise.	He/she experiences having limited insight into his/her state of health, with limited knowledge of how to deal with health problems when they arise.	He/she experiences having good insight into his/her state of health, with sufficient knowledge of how to deal with health problems when they arise.
heiQ4: constructive attitudes and approaches	He/she finds that his/her health problems limit him/her from enjoying life.	He/she finds that his/her health problems limit him/her from enjoying life to some extent.	He/she does not find that his/her health problems limit him/her from enjoying life.
heiQ5: skill and technique acquisition	He/she experiences having insufficient skills and techniques to deal with health problems when they arise.	He/she experiences having limited skills and techniques to deal with health problems as they arise.	He/she experiences having sufficient skills and techniques to deal with health problems as they arise.
heiQ8: emotional distress	He/she generally experiences a high degree of emotional distress due to his/her health.	He/she generally experiences a bit of emotional distress due to his/her health.	He/she generally experiences no emotional distress due to his/her health.
HLQ4[b]: social support for health	He/she does not feel understood or supported by his/her social network.	He/she feels only partially understood and supported by his/her social network.	He/she feels adequately understood and supported by his/her social network.
HLQ1: feel understood and supported by health care providers	He/she experiences a lack of understanding and support from the health care professionals he/she has contact with and access to.	He/she experiences limited understanding and support from the health care professionals he/she has contact with and access to.	He/she experiences adequate understanding and support from the health care professionals she/he has contact with and access to.
eHLQ2[c]: understanding of health concepts and language	He/she experiences having trouble understanding information about health and illness and may experience insufficient knowledge to have conversations with others about this.	He/she may experience problems understanding information about health and illness and may experience having limited knowledge to have conversations with others about this.	He/she does not experience having problems understanding information about health and illness and has enough knowledge to have conversations with others about this.
eHLQ1: using technology to process health information	He/she finds that he/she has have insufficient knowledge about how he/she can use technology to take care of his/her health.	He/she finds that he/she has limited knowledge about how he/she can use technology to take care of his/her health.	He/she has sufficient knowledge about how he/she can use technology to take care of his/her health.
eHLQ3: ability to actively engage with digital services	He/she has insufficient knowledge of how to use technology to navigate health care systems.	He/she has limited knowledge of how to use technology to navigate health systems.	He/she has good knowledge of how to use technology and can navigate health systems easily.
eHLQ6: access to digital services that work	He/she experiences insufficient access to digital health systems that work and can be accessed by himself/herself and others who need it.	He/she experiences having limited access to digital health systems that work and can be accessed by himself/herself and others who need it.	He/she experiences having sufficient access to digital health systems that work and can be accessed by himself/herself and others who need it.
eHLQ7: digital services that suit individual needs	He/she believes that digital services are insufficiently able to adapt to his/her needs.	He/she believes that digital services can adapt to his/her needs to a limited extent.	He/she believes that digital services can adequately adapt to his/her needs.
eHLQ5: motivated to engage with digital services	He/she finds that technology is insufficiently helpful and helps him/her keep up with and take care of his/her health.	He/she finds that technology is helpful to a limited extent and helps him/her keep up with and take care of his/her health.	He/she finds that technology is sufficiently useful and helps them keep up with and take care of his/her health.
eHLQ4: feel safe and in control	He/she feels uneasy about how his/her health data are used by others.	He/she feels partly uncomfortable with how his/her health data are used by others.	He/she feels comfortable with how his/her health data are used by others.

[a]heiQ: Health Education Impact Questionnaire.

[b]HLQ: Health Literacy Questionnaire.

[c]eHLQ: eHealth Literacy Questionnaire.

Data Sampling and Analysis

Data collection involved conducting cognitive interviews with each participant to validate the individual text vignettes developed from the individuals' READHY responses. These interviews allowed for iterative feedback and refinement of the text string to ensure their coherence and relevance. The analysis focused on identifying and addressing any issues in understanding and meaningfulness of the text strings and vignettes.

Iterative Validation of Data-Informed Text Vignettes

The text strings were validated through cognitive interviews with individuals and clinicians in an iterative process. Cognitive

interviews, often used for validating questionnaires, involve participants "thinking out loud" by verbally reporting their mental activity while engaging with an item in a questionnaire [34,35]. These interviews are conducted using a semistructured approach, with in-depth sessions where participants read a text vignette aloud and provide feedback by sharing their thoughts and insights throughout the process.

The interviews were conducted in 2 stages: first with individuals and then with clinicians. Before each individual's interview, they completed READHY, which was then used to create individual text vignettes. Each individual was presented with a full vignette, created based on their previous responses, to gather feedback on both individual text strings and the vignette as a coherent whole, evaluating its clarity and personal relevance.

After each interview, the data were analyzed individually to identify issues in the understanding and meaningfulness of the text vignettes. These individual findings were then aggregated across interviews to identify recurring problems, leading to adjustments of the text vignettes.

The iterative process included multiple stages of feedback incorporation. The first iteration considered feedback from 3 participants, with subsequent iterations incorporating feedback from additional participants—3 in the second and third iterations and 2 in the fourth.

Identified problems were categorized informally into 4 groups: scale-specific wording changes, need for further specification, issues with text vignette setup, and difficulties in answering the instrument [34].

Once all individual interviews were completed and the text vignettes and algorithm were adjusted, clinicians were introduced to the revised vignettes. The assignment of vignettes to clinicians was random, as they had no prior knowledge of or relationship with the individuals. Clinicians provided feedback on these vignettes, which led to further refinements. Each revised version was then presented to subsequent clinicians for additional input, ensuring that the final text was coherent and well understood.

Part 3 Methods

Data Sampling

First, 2 patients completed the instrument. From their responses, 2 individual text vignettes were constructed using the same method as applied in part 2. The average scale scores were calculated for each scale, and the scale score was then matched with the final adjusted corresponding sentence (Table 2). Afterward, the 13 text strings were combined to create an individual text vignette. These text vignettes were then presented to the 3 clinicians in 3 individual semistructured interviews designed to explore the text vignettes within an actual clinical environment and gain the clinicians' perspective. Each clinician was presented with 1 text vignette. Because all clinicians were somewhat familiar with booth patients, the specific vignette used in the interviews was randomly chosen based on which vignette was completed at the time of the interview.

Patients were also interviewed individually to assess their recognition of themselves in their text vignettes and whether it made sense for them in a clinical environment. However, the primary focus in part 3 was on the clinicians. Patient interviews were conducted via phone at the request of both patients. The interviews were recorded with consent and later transcribed.

The unique aspect lies in its exploration of text vignettes within clinical contexts, focusing on clinicians who interact with the patients described, and incorporating insights from their practical experience.

Analysis

Two interview guides were prepared, 1 for patient interviews and 1 for clinician interviews (Multimedia Appendix 1). These questions were grounded in the central inquiry: How can text vignettes generated by PRO data, as an alternative or supplement to traditional quantitative representations, improve dialogue and understanding between clinicians and patients regarding PROs?

All interviews were analyzed using qualitative content analysis using inductive coding to identify and describe patterns, categories, and meanings within the data [36]. The analysis was conducted using NVivo software (version 12.4.0; Lumivero) [37]. From the coding process, 4 main categories emerged: "recognition and confirmation," "challenges with context and conceptual understanding," "one size does not fit all," and "reflections on using text vignettes."

An example of a text vignette, as presented to both clinicians and patients, is shown in Textbox 2 (translated by the authors). This vignette is based on one of the patient's responses to the READHY instrument.

Textbox 2. Text vignette based on Readiness and Enablement Index for Health Technology.

She experiences having limited insight into her state of health, with limited knowledge of how to deal with health problems when they arise.

She finds that her health issues to some extent limit her from enjoying life.

She experiences having sufficient skills and techniques to deal with health problems as they arise and generally, experiences no emotional distress due to her health.

She feels adequately understood and supported by her social network and experiences adequate understanding and support from the healthcare professionals she has contact and access to.

She experiences having trouble understanding information about health and illness and may experience insufficient knowledge to have a conversation with others about this.

She finds that she has insufficient knowledge about how she can use technology to take care of her health and has insufficient knowledge of how to use technology to navigate healthcare systems.

She experiences having sufficient access to digital health systems that work and can be accessed by herself and others who need it, and that digital services can adequately adapt to her needs.

She finds that technology is helpful to a limited extent and helps her keep up with and take care of her health.

She feels comfortable with how her health data is used by others.

Results

Part 1

Four clusters with various levels of high, intermediate, and low average scores of scales within self-management, social capital, and digital health literacy were created (Table 3).

The mean value for each scale was presented for each of the 4 clusters that were informed. Values considered to represent a problematic level were indicated in red (<2.00), values considered to be below a sufficient level were indicated in yellow (2.00-2.50), and values considered to represent a sufficient level were represented in green (>2.50).

Table 3. Stratification using the Readiness and Enablement Index for Health Technology.

Metric (mean values)	Cluster 1 (n=16)	Cluster 2 (n=5)	Cluster 3 (n=11)	Cluster 4 (n=9)
heiQ3[a]: self-monitoring and insight	2.98	3.66	3.13	3.11
heiQ4: constructive attitudes and approaches	2.81	3.28	2.76	3.46
heiQ5: skill and technique acquisition	2.65	3.15	2.84	3.11
heiQ8: emotional distress	2.50	2.76	2.13	3.12
HLQ1[b]: feel understood and supported by health care providers	2.78	3.70	3.40	3.44
HLQ3: actively managing my health	2.68	3.40	2.92	2.68
HLQ4: social support for health	2.83	3.59	2.90	3.33
eHLQ1[c]: using technology to process health information	2.78	3.92	1.89	2.03
eHLQ2: understanding of health concepts and language	2.85	3.96	2.82	2.77
eHLQ3: ability to actively engage with digital services	2.95	3.96	1.99	2.65
eHLQ4: feel safe and in control	2.97	4.00	2.95	3.18
eHLQ5: motivated to engage with digital services	2.85	3.92	2.11	2.60
eHLQ6: access to digital services that work	2.86	3.96	2.42	2.97
eHLQ7: digital services that suit individual needs	2.78	3.80	2.04	2.68

[a]heiQ: Health Education Impact Questionnaire.

[b]HLQ: Health Literacy Questionnaire.

[c]eHLQ: eHealth Literacy Questionnaire.

On the basis of the thresholds, a focused text vignette was constructed almost as a narrative for each cluster together with key sociodemographic characteristics to create 4 text vignettes, wherein each represented an archetype of the profile—termed a "persona" (Multimedia Appendix 2). To simplify the information in this initial phase, only scales with high (initial green color) or low (initial red) average scores (>2.50 or <2.0) were included in the text vignettes, thereby creating focused text vignettes highlighting potential areas of barriers and resources.

EKW and LK then presented these text vignettes to colleagues in the Smart Inclusive Living Environments project. On the basis of the insights gained from the proof of concept with text vignettes at group or personal level, it was decided to proceed to part 2, focusing on the development of structured text strings that encompass high, intermediate, and low average score levels.

Part 2

A total of 39 initial text strings were created, with 3 strings for each of the 13 READHY scales, reflecting different levels of knowledge, skills, or attitudes within each scale. These text strings were designed to align with the scale names, which summarize the relevant experiences. The initial text strings are included in Multimedia Appendix 3.

Immediate feedback from the first 3 participants indicated that they could easily recognize themselves in the text vignettes, found them meaningful, and saw how their responses were reflected in them.

In the first iteration of text vignette validation, participants largely recognized themselves but identified issues related to the structure of the text vignettes and the completion of the instrument. Feedback included suggestions to revise lengthy sentences and reduce academic language, with specific terms

such as “health professionals” and “health-related concepts” being difficult to understand. Participants also struggled with the term “technology” used in the text vignette, such as in the description of scale eHLQ3.

On the basis of cognitive interviews, “healthcare providers” was changed to “healthcare professionals” and “health-related concepts” was simplified to “information of health and illness.” The order in which the scales were presented was also slightly changed. The order of HLQ4 and HLQ1 was adjusted to create a more natural flow when reading through the vignette. Similarly, the eHLQ items were reordered to enhance the coherence of the vignette.

In subsequent validation iterations, feedback was generally positive, with individuals finding the text vignettes understandable and recognizable. However, they struggled more with areas where they rated themselves as limited or insufficient. Some participants mentioned difficulties with the questionnaire due to the lack of a “don’t know” option and noted that terms such as “insufficient” felt too academic and unrepresentative of their self-view. The term “digital services” was also revised to “digital health services” to clarify its meaning. Participants suggested that discrepancies in text vignette descriptions might stem from specific episodes rather than their general experiences, and some participants requested more precise wording to better reflect their experiences. During iterations and refinement based on feedback, the understandability and recognizability of the vignettes improved, with individuals having an overall positive view of the text vignettes after final refinement.

The clinicians had a generally positive view about the generated READHY text vignettes. Feedback emphasized the need for clarity in highlighting that the text vignettes were based on a PRO instrument. The length of the text vignettes had mixed reviews. Feedback suggested categorizing content into taxonomic groups. The language was too academic, and adjustments had to be made to align with the instrument.

Clinicians see the value of using the text vignettes at the population level, allowing for generalization and description of specific populations based on competencies and characteristics. However, the current format may not be suitable for individual-level use, as it offers a simplified representation for personalized interaction. On the basis of input received from all participants, the strings were refined. The refined text strings are presented in Table 3 (translated from Danish to English by the authors). In the context of this study, we have used gender pronouns, but they should be adjusted to ensure relevance and significance within the context and setting.

Part 3

There was a positive attitude among the clinicians and patients toward the text vignette format and this way of presenting and communicating PROs. The clinicians experienced being able to recognize patients in the text vignettes, and patients experienced being able to recognize themselves in their text vignettes. Although the text vignette format does not necessarily meet the specific needs, the textual presentation makes the information from PROs more accessible in an understandable way for both clinicians and patients.

There are different preferences for how dynamic the tool should be. Some prefer to select only specific dimensions, while others believe that the strength of the text vignettes comes from them providing the full picture insights and exposing possibly overlooked aspects. Clinicians who favor the text vignette suggest that color coding as a supplement to the text vignette would be helpful and that a schematic summary would help with readability, indicating that flexibility in the profile presentation is needed to accommodate different preferences. Some prefer text-based representations, while others find visual presentations more intuitive.

Clinicians may face language and comprehension challenges due to ambiguity in words and abstract concepts within the text vignettes. Ambiguity in words and concepts, such as “to some extent,” leaves clinicians uncertain about the severity of the patient’s problems. Abstract concepts such as “health” and “technology” also pose challenges in interpretation. The text vignettes highlight new angles in patients’ self-perception, health, motivation, and emotional distress, which clinicians find crucial.

Opinions differ on whether to highlight strengths or weaknesses. Some prefer insights into both aspects, while others prefer focusing solely on the weaknesses that are most likely to be approved. Clinicians acknowledge the value of using text vignettes at a population or group level for generalization but point out some limitations for individual-level use due to simplified representations. Emphasis on personal dialogue with patients remains crucial for addressing unresolved questions and determining personalized interactions.

From the patient’s viewpoint, there is self-recognition and understanding; however, engaging in discussions about text vignette representation proves challenging. It is essential to note that clinicians report based on their experience of interacting with patients, highlighting the dynamic nature of this interaction in understanding and interpreting the text vignettes. The clinician’s insights provide a valuable perspective for understanding the broader implications of text vignettes in a clinical setting.

Discussion

Principal Findings

This study presents a process for developing text vignettes based on PRO data, resulting in the creation of individual vignettes that incorporate all 13 scales of the READHY framework. The motivation for the development of these text vignettes was the challenges associated with presenting PRO data as numeric values or graphical representations, aiming to provide a clearer and more engaging way to present and communicate PRO information.

This findings of this study suggest an overall positive attitude toward the PRO data–informed text vignette format and this way of presenting and communicating PROs. Clinicians experienced being able to recognize patients in the text vignettes,

and the patient's experienced being able to recognize themselves in and understand the text vignettes describing themselves.

Considering the background emphasizing the challenges associated with graphical representations of PRO data, the findings suggest that PRO-informed text vignettes are understandable for clinicians and patients. While the text format may not perfectly align with all the needs and interests of clinicians and patients, it presents PRO data in a comprehendible manner, with the text vignettes presenting a recognizable description.

Patients' Perspective

Patients found elements of self-recognition in the text vignettes, although discussing their representation could be challenging. It is important to note that text vignettes also face challenges related to health literacy and differences in understanding, as highlighted in part 3. These challenges can pose significant barriers for individuals with low health literacy or those who cannot read. In such cases, regardless of the presentation format, additional support, such as reading the vignettes aloud, may be necessary. However, the written format generally makes the data more accessible and less complex.

Another potential challenge is that the threshold values used to select appropriate text strings may result in choices that are not recognized by the informants. For example, individuals scoring between 2.00 and 2.70 might not feel limited, or they may perceive issues in areas not fully captured by the items in a scale. This issue requires further exploration in future studies to determine whether scales can be used effectively or whether a focus on single items is more appropriate, particularly for those scoring at the extremes. This may be necessary as we work in the intersection of presenting data from psychometric valid scales reporting on latent variables and presenting numeric values as an interpretation of experiences. The patients' experiences and feelings may sometimes be undermined or misinterpreted in this process.

Clinicians' Perspective

The data-informed text vignettes add a candid dimension to the information, making it directly engaging and allowing for the capture of new perspectives on patients that clinicians would not otherwise have focused on, although not all perspectives are equally relevant. It is especially the aspects of experienced support and patients' own knowledge about health and illness that are important for clinicians to understand and gain perspective on. There are different preferences among clinicians regarding the format of the text vignettes; some prefer text-based representations, while others prefer visual presentations. Likewise, there are different preferences regarding how dynamic the text-based tool should be. Some experience a need to be able to select only specific dimensions, while others believe it is part of the strength of the vignettes that it presents the full picture and thus insight into possibly overlooked aspects. This indicates that there is a need for flexibility in the text vignette presentation to accommodate different preferences.

Clinical Relevance

Clinicians using text vignettes stress the importance of technology that effectively addresses practical and treatment-specific needs. It remains imperative that the information within text vignettes aligns with the treatment being offered.

Clinicians emphasize that the central focus should be on understanding the condition and enabling patients to lead fulfilling lives. Reading the text vignettes prompts clinicians to reflect on aspects that may have been previously overlooked, serving as a reminder of the critical need for clinical relevance in any work involving PROs. It underlines that for PROs, and indeed text vignettes, to realize its full potential in creating value, it must align with the clinical context and be relevant to clinical practice.

A More Straightforward Representation of PROs

In addressing some of the challenges with the quantitative representation of PROs, the use of text vignettes allows for a deeper understanding of PROs. Text vignettes recognize the complexities of patients' experiences and provide a broader perspective. This approach prevents reducing patients to just numerical data, offering instead a more comprehensive view of an individual's health. The text vignettes are presented in an understandable and consistent format, with transparency about how the collected information is applied. This transparency is crucial for patients' effective use of PROs [38]. By transforming complicated numerical data into a clearer and more coherent text format, text vignettes offer a direct representation that ensures patients and clinicians can grasp details without relying on prior knowledge of graphical presentations. This textual format encourages conversation, minimizes ambiguity, and fosters dialogue for a better understanding of individual experiences.

Promoting Dialogue for Mutual Understanding

A data-informed text vignette is not a universal solution, but rather an approach with both advantages and disadvantages that must be weighted and adapted to individual preferences and needs and that requires a more dynamic approach to meet needs. Recognizing these challenges suggests a need for meaningful conversations between patients and clinicians about aligning their understanding at a conceptual level. Such discussions offer a chance to clarify word meanings and ensure a common understanding of the patient's experiences. This approach is essential for accurate and meaningful communication between patients and clinicians, ultimately improving the quality of care and treatment. Clinicians and patients may have different perspectives on what the patient reports in the PRO instrument. An example is how HLQ scores were not perceived in the same way by patients and their physicians, with a disconcordance ranging between 20% and 44% in the 9 HLQ scales [14]. Translating the data into text vignettes may ease the conversation and help identify these disconcordances for other PRO data also, ease the alignment of expectations, and provide a common understanding of the patient perspective.

Bias Concern

Clinicians are generally receptive to text vignettes and the insights they provide. Many clinicians feel that the text vignettes offer a strong sense of the patients before any direct contact is made. However, clinicians also express concerns that the text vignette may make them biased in their encounter with the patient, which may affect their ability to understand and treat the patient objectively. It emphasizes the necessity of acquiring proficiency in tool use, particularly when presenting information in a text format. Being mindful of potential biases becomes central, especially in a context where the text vignette format might intensify the impact. This highlights the importance of a nuanced and skillful application of the tool, recognizing that biases are not exclusive to text vignette representation but should be carefully navigated, especially when conveyed in a confrontational textual manner.

Strengths and Limitations

A strength of this study lies in the stepwise process of creating a dataset using a validated instrument, followed by data triangulation. This process is based on iteratively developed text lines that can be combined into cohesive text vignettes, forming complete paragraphs. The cocreation with individuals, people living with a chronic condition, municipalities, hospital-based clinicians, and 1 general practitioner helps ensure the validity and meaningfulness of the proposed solution.

The number of participants in each of the 3 parts may be considered a limitation. In part 1, convenience sampling was used among participants active in a telehealth service and was limited by availability. We do not find this to be a problem for the presented results, as we were able to create 4 profiles and personas that are meaningful and can serve to illustrate the principles used as background for parts 2 and 3. The number of participants in part 2 was determined through concurrent evaluation during data sampling. When the inclusion of additional informants no longer contributed new information, the sampling process was considered complete. By including data from 2 distinct sociodemographic areas, we aimed to capture a range of perspectives. On the basis of these arguments, we found the number to have sufficient information power for a proof-of-concept study [39]. This is also supported by most of the participants finding the text vignettes understandable and recognizable.

In particular, in part 3, the number of patients is likely to be too small, but we have included this part to illustrate how clinicians who are familiar with the patients respond to the text vignettes describing them. The findings need to be assessed in this light. Moreover, the results may differ from those that will be obtained using PRO data instruments that are developed for clinical purposes, such as European Organisation For Research And Treatment Of Cancer [40], The World Health Organization-Five Well-Being Index [41], or European Quality of life–5 Dimensions [42].

Perspective

The study has used the READHY instrument to demonstrate how data-based text vignettes can help to understand patients and align expectations about what the data tell. During the field studies and in discussions afterward, we have received expressions of interest in setting up algorithms to convert other established PRO data results into text vignettes. This calls for abundant work in this field in collaboration with several PRO instrument owners and distributors, and we welcome collaboration in this field. It also calls for atomization programming interfaces that build on the proposed structures with filters based on cutoff values combined with generative artificial intelligence to refine the created text vignettes.

Conclusions

The study demonstrates how text vignettes can be developed using PRO instruments, with individual scales as input strings. This provides an opportunity to present numeric values in a text format that is understandable and recognizable to most patients and clinicians. Some clinicians prefer text vignettes to other forms of presentations and find them especially useful in establishing a better understanding of a patient before initial contact. However, language and the arrangement of text vignettes can present challenges, underscoring the need for adaptation to specific needs and clinician training to ensure accurate interpretation of the text vignettes. Different preferences for text vignette formats highlight the need for a more dynamic and adaptable tool that can cater to individual needs and preferences.

Acknowledgments

The authors would like to thank all participants for their contribution. The authors would like to specially thank Torsten Risør, Sine Rossen, Rikke Terp, Thyge Lynghøj Nielsen, and the PreCareKlinikken for providing feedback throughout the process and Mette Krog Andreasen and Morten Brandrup for facilitating access to informants.

The authors thank the European Union (EU) Horizon 2020 and all partners in Smart Inclusive Living Environments (SMILE) for making this work possible. This research was funded by the EU Horizon 2020 Research and Innovation Program under grant 101016848.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guides.

[DOCX File , 17 KB - humanfactors_v12i1e58077_app1.docx]

Multimedia Appendix 2
Group-based text vignettes.
[DOCX File , 640 KB - humanfactors_v12i1e58077_app2.docx]

Multimedia Appendix 3
Initial text strings based on scores.
[PDF File (Adobe PDF File), 63 KB - humanfactors_v12i1e58077_app3.pdf]

References

1. Anvendelse af PRO-data. Danske Patienter. URL: https://danskepatienter.dk/vibis/inddragelse-i-eget-forloeb/pro-data/anvendelse-af-pro-data [accessed 2023-02-02]
2. Sandø N. Sundhedskompetence i det danske sundhedsvæsen. Copenhagen, Demark: Sundhedsstyrelsen; Aug 10, 2022.
3. Sundheds kompetence i det danske sundhedsvæsen - En vej til mere lighed. Sundhedsstyrelsen. URL: https://www.sst.dk/da/viden/Sundhedsvaesen/Ulighed-i-sundhed/Sundhedskompetence [accessed 2023-03-22]
4. Meadows K. Do patient-reported outcome measures tell us the full story? Clin Nurs Res 2022 Feb;31(2):159-162. [doi: 10.1177/10547738221078335] [Medline: 35125032]
5. Kwon JY, Russell L, Coles T, Klaassen RJ, Schick-Makaroff K, Sibley KM, et al. Patient-reported outcomes measurement in radiation oncology: interpretation of individual scores and change over time in clinical practice. Curr Oncol 2022 Apr 27;29(5):3093-3103 [FREE Full text] [doi: 10.3390/curroncol29050251] [Medline: 35621641]
6. Albers EA, Fraterman I, Walraven I, Wilthagen E, Schagen SB, van der Ploeg IM, et al. Visualization formats of patient-reported outcome measures in clinical practice: a systematic review about preferences and interpretation accuracy. J Patient Rep Outcomes 2022 Mar 03;6(1):18 [FREE Full text] [doi: 10.1186/s41687-022-00424-3] [Medline: 35239055]
7. Grossman LV, Feiner SK, Mitchell EG, Masterson Creber RM. Leveraging patient-reported outcomes using data visualization. Appl Clin Inform 2018 Jul;9(3):565-575 [FREE Full text] [doi: 10.1055/s-0038-1667041] [Medline: 30068012]
8. Liu LH, Garrett SB, Li J, Ragouzeos D, Berrean B, Dohan D, et al. Patient and clinician perspectives on a patient-facing dashboard that visualizes patient reported outcomes in rheumatoid arthritis. Health Expect 2020 Aug;23(4):846-859 [FREE Full text] [doi: 10.1111/hex.13057] [Medline: 32270591]
9. Kuijpers W, Giesinger JM, Zabernigg A, Young T, Friend E, Tomaszewska IM, et al. Patients' and health professionals' understanding of and preferences for graphical presentation styles for individual-level EORTC QLQ-C30 scores. Qual Life Res 2016 Mar;25(3):595-604 [FREE Full text] [doi: 10.1007/s11136-015-1107-3] [Medline: 26353905]
10. Redegørelse Om Ældre Området 2020. Sundheds- og Ældreministeriet. 2020. URL: https://aeldremin.dk/Media/637666126708148318/AE_Redegoerelse_om_aeldreomraadet_2020.pdf [accessed 2024-04-29]
11. Loth FL, Holzner B, Sztankay M, Bliem HR, Raoufi S, Rumpold G, et al. Cancer patients' understanding of longitudinal EORTC QLQ-C30 scores presented as bar charts. Patient Educ Couns 2016 Dec;99(12):2012-2017 [FREE Full text] [doi: 10.1016/j.pec.2016.08.004] [Medline: 27506581]
12. Campbell R, Ju A, King MT, Rutherford C. Perceived benefits and limitations of using patient-reported outcome measures in clinical practice with individual patients: a systematic review of qualitative studies. Qual Life Res 2022 Jun;31(6):1597-1620. [doi: 10.1007/s11136-021-03003-z] [Medline: 34580822]
13. Munzner T. Visualization Analysis and Design: Opportunities for Domestic Investments in Water and Sanitation for the Poor. Boca Raton, FL: CRC Press; 2015.
14. Hawkins M, Gill SD, Batterham R, Elsworth GR, Osborne RH. The Health Literacy Questionnaire (HLQ) at the patient-clinician interface: a qualitative study of what patients and clinicians mean by their HLQ scores. BMC Health Serv Res 2017 Apr 27;17(1):309 [FREE Full text] [doi: 10.1186/s12913-017-2254-8] [Medline: 28449680]
15. El-Daly I, Ibraheim H, Rajakulendran K, Culpan P, Bates P. Are patient-reported outcome measures in orthopaedics easily read by patients? Clin Orthop Relat Res 2016 Jan;474(1):246-255 [FREE Full text] [doi: 10.1007/s11999-015-4595-0] [Medline: 26472587]
16. Beauchamp A, Dodson S, Batterham R, Asproloupos D, Richard OC, Cavalli S, et al. Ophelia manual how to apply the Ophelia health literacy process for improving health outcomes. Swinburne University of Technology. 2017. URL: https://www.nswmentalhealthcommission.com.au/sites/default/files/2021-06/swinburne-_ophelia_manual_orange.pdf [accessed 2024-04-29]
17. Batterham RW, Hawkins M, Collins PA, Buchbinder R, Osborne RH. Health literacy: applying current concepts to improve health services and reduce health inequalities. Public Health 2016 Mar;132:3-12. [doi: 10.1016/j.puhe.2016.01.001] [Medline: 26872738]
18. Bakker MM, Putrik P, Aaby A, Debussche X, Morrissey J, Råheim BC, et al. Acting together – WHO national health literacy demonstration projects (NHLDPs) address health literacy needs in the European region. World Health Organization.

2019. URL: https://iris.who.int/bitstream/handle/10665/327059/php-5-2-3-233-243-eng.pdf?sequence=1&isAllowed=y [accessed 2024-04-29]
19. Kayser L, Rossen S, Karnoe A, Elsworth G, Vibe-Petersen J, Christensen JF, et al. Development of the multidimensional readiness and enablement index for health technology (READHY) tool to measure individuals' health technology readiness: initial testing in a cancer rehabilitation setting. J Med Internet Res 2019 Feb 12;21(2):e10377 [FREE Full text] [doi: 10.2196/10377] [Medline: 30747717]
20. Kayser L, Karnoe A, Duminski E, Jakobsen S, Terp R, Dansholm S, et al. Health professionals' eHealth literacy and system experience before and 3 months after the implementation of an electronic health record system: longitudinal study. JMIR Hum Factors 2022 Apr 29;9(2):e29780 [FREE Full text] [doi: 10.2196/29780] [Medline: 35486414]
21. Mankins JC. Technology readiness assessments: a retrospective. Acta Astronautica 2009 Nov;65(9-10):1216-1223. [doi: 10.1016/j.actaastro.2009.03.058]
22. Hvad er PreCareKlinikken? PreCareKlinikken. URL: https://precareklinikken.dk/om/ [accessed 2023-05-21]
23. Phanareth K, Dam AL, Hansen MA, Lindskrog S, Vingtoft S, Kayser L. Revealing the nature of chronic obstructive pulmonary disease using self-tracking and analysis of contact patterns: longitudinal study. J Med Internet Res 2021 Oct 19;23(10):e22567 [FREE Full text] [doi: 10.2196/22567] [Medline: 34665151]
24. Home page. SMILE eHealth. URL: https://smileehealth.eu/home [accessed 2024-01-13]
25. International standard classification of education ISCED 2011. UNESCO Institute for Statistics. URL: https://uis.unesco.org/en/topic/international-standard-classification-education-isced [accessed 2024-04-29]
26. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. Acta Oncol 2019 May;58(5):610-618. [doi: 10.1080/0284186X.2018.1562213] [Medline: 30698060]
27. Terp R, Kayser L, Lindhardt T. Older patients' competence, preferences, and attitudes toward digital technology use: explorative study. JMIR Hum Factors 2021 May 14;8(2):e27005 [FREE Full text] [doi: 10.2196/27005] [Medline: 33988512]
28. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): an outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
29. Sørensen K, Van den Broucke S, Pelikan JM, Fullam J, Doyle G, Slonska Z, HLS-EU Consortium. Measuring health literacy in populations: illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). BMC Public Health 2013 Oct 10;13(1):948 [FREE Full text] [doi: 10.1186/1471-2458-13-948] [Medline: 24112855]
30. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth literacy questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
31. Overview of mandatory reporting | Danish research ethics committees. The Danish National Center for Ethics. URL: https://researchethics.dk/information-for-researchers/overview-of-mandatory-reporting [accessed 2024-03-05]
32. IBM SPSS statistics for Windows. IBM Corp. URL: https://www.ibm.com/products/spss-statistics [accessed 2024-04-29]
33. Poulsen HS, Eiriksson SD, Christiansen AS, Wingstrand A. Sundhedsprofil 2021 for Region Sjælland og kommuner: »Hvordan har du det?«. Region Sjælland. URL: https://pure-portal.regsj.dk/da/publications/sundhedsprofil-2021-for-region-sj%C3%A6lland-og-kommuner-hvordan-har-d [accessed 2024-04-29]
34. Willis GB. Analyzing and documenting cognitive interview results. In: Willis GB, editor. Cognitive Interviewing. Thousand Oaks, CA: Sage Publications; 2011.
35. Conrad F, Blair J. From impressions to data: Increasing the objectivity of cognitive interviews. ResearchGate. 1996. URL: https://www.researchgate.net/publication/237601397_From_impressions_to_data_Increasing_the_objectivity_of_cognitive_interviews [accessed 2024-04-29]
36. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
37. NVivo. Lumivero. URL: https://lumivero.com/products/nvivo/ [accessed 2024-04-29]
38. Carfora L, Foley CM, Hagi-Diakou P, Lesty PJ, Sandstrom ML, Ramsey I, et al. Patients' experiences and perspectives of patient-reported outcome measures in clinical care: a systematic review and qualitative meta-synthesis. PLoS One 2022 Apr 21;17(4):e0267030 [FREE Full text] [doi: 10.1371/journal.pone.0267030] [Medline: 35446885]
39. Malterud K, Siersma VD, Guassora AD. Sample size in qualitative interview studies: guided by information power. Qual Health Res 2016 Nov;26(13):1753-1760. [doi: 10.1177/1049732315617444] [Medline: 26613970]
40. Wintner LM, Sztankay M, Aaronson N, Bottomley A, Giesinger JM, Groenvold M, EORTC Quality of Life Group. The use of EORTC measures in daily clinical practice-a synopsis of a newly developed manual. Eur J Cancer 2016 Nov;68:73-81 [FREE Full text] [doi: 10.1016/j.ejca.2016.08.024] [Medline: 27721057]
41. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 well-being index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
42. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]

Abbreviations

COPD: chronic obstructive pulmonary disease
eHLQ: eHealth Literacy Questionnaire
heiQ: Health Education Impact Questionnaire
HLQ: Health Literacy Questionnaire
OPHELIA: Optimizing Health Literacy and Access
PRO: patient-reported outcome
READHY: Readiness and Enablement Index for Health Technology

Edited by A Kushniruk, D Kaufman; submitted 05.03.24; peer-reviewed by J Yang, C Cheng, Y Chu; comments to author 04.08.24; revised version received 04.09.24; accepted 31.12.24; published 05.02.25.

Please cite as:
Kelly SC, Wegener EK, Kayser L
Creation of Text Vignettes Based on Patient-Reported Data to Facilitate a Better Understanding of the Patient Perspective: Design Study
JMIR Hum Factors 2025;12:e58077
URL: https://humanfactors.jmir.org/2025/1/e58077
doi:10.2196/58077
PMID:

Original Paper

Development of a Web-Based Intervention for Middle Managers to Enhance Resilience at the Individual, Team, and Organizational Levels in Health Care Systems: Multiphase Study

Eva Gil-Hernández[1], MSc; Irene Carrillo[2], PhD; Jimmy Martin-Delgado[3,4], MD, PhD; Daniel García-Torres[1], MSc; José Joaquín Mira[1,2,5], MPH, PhD

[1]ATENEA Research Group, Foundation for the Promotion of Health and Biomedical Research of Valencia Region (FISABIO), Alicante, Spain

[2]Health Psychology Department, Universidad Miguel Hernández, Elche, Spain

[3]Junta de Beneficencia de Guayaquil, Hospital de Especialidades Alfredo Paulson, Portoviejo, Ecuador

[4]Instituto de Investigación e Innovación en Salud Integral, Facultad de Ciencias de la Salud, Universidad Católica de Santiago de Guayaquil, Guayaquil, Ecuador

[5]Conselleria de Sanidad, Health Centre Hospital Provincial-Pla, Alicante-Sant Joan Health District, Alicante, Spain

Corresponding Author:
José Joaquín Mira, MPH, PhD
Health Psychology Department
Universidad Miguel Hernández
Avinguda de la Universitat d'Elx
Elche, 03202
Spain
Phone: 34 966658984
Email: jose.mira@umh.es

Abstract

Background: Health care institutions face high systemic risk due to the inherent uncertainty and complexity of their operations. This often leads to stressful incidents impacting the well-being of health care professionals, which can compromise the effectiveness of health care systems. Enhancing resilience among health care professionals is essential for maintaining high-quality care and ensuring patient safety. The role of middle managers is essential to ensure the response capacity of individuals and teams.

Objective: This study aims to develop a web-based intervention aimed at middle management to enhance individual, team, and organizational resilience.

Methods: An observational study was conducted in 3 phases: design, validation, and pilot study. The study was initiated in February 2022 and concluded in June 2023. Phase 1 involved designing the content for the web-based tool based on a comprehensive review of critical elements around resilience. Phase 2 included validation by an international panel of experts who reviewed the tool and rated it according to a structured grid. They were also encouraged to highlight strengths and areas for improvement. Phase 3 involved piloting the tool with health care professionals in Ecuador to refine the platform and assess its effectiveness. A total of 458 people were invited to participate through the Institutional Course on Continuous Improvement in Health Care Quality and Safety offered by the Ministry of Public Health of Ecuador.

Results: The tool, eResiliencia, was structured into 2 main blocks: individual and team resilience and organizational resilience. It included videos, images, PDFs, and links to dynamic graphics and additional texts. Furthermore, 13 (65%) of the 20 experts validated the tool, rating content clarity at an average of 4.5 (SD 0.7) and utility at an average of 4.7 (SD 0.5) out of 5. The average overall satisfaction was 9.3 (SD 0.6) out of 10 points, and feedback on improvements was implemented. A total of 362 health care professionals began the intervention, of which 218 (60.2%) completed preintervention and postintervention questionnaires, with significant knowledge increases (P<.001). Of the 362 health care professionals, 146 (40.3%) completed the satisfaction questionnaire, where overall satisfaction was rated at an average of 9.4 (SD 1.1) out of 10 points.

Conclusions: The eResiliencia web-based platform provides middle managers with resources to enhance resilience among their teams and their components, promoting better well-being and performance, even under highly stressful events. Future research should focus on long-term impacts and practical applications in diverse clinical settings.

(JMIR Hum Factors 2025;12:e67263) doi:10.2196/67263

KEYWORDS

resilience; health care professionals; web-based intervention; middle management; well-being; patient safety

Introduction

Systemic Risk

Health care institutions are considered high-reliability organizations due to the inherent systemic risk associated with the activities they conduct [1]. Uncertainty and complexity, inherent in health care delivery along with other factors, contribute to the occurrence of higher stressful incidents that sometimes result in harm (adverse events) [2]. Health care professionals are exposed almost daily to stressful situations that can deteriorate their occupational well-being, thus compromising the effectiveness of health care systems [3]. In addition, the impact and demands stemming from unexpected health outcomes and health care crises further exacerbate this situation. Ultimately, the post–COVID-19 pandemic period has underscored a reality that has always existed but has not been adequately addressed until now: the well-being of health care professionals is a necessary condition to ensure the effectiveness of health care systems and the quality of care [4].

Professional Well-Being

In clinical contexts, situations with the potential to deteriorate the well-being of professionals are varied. These situations can range from those related to patient safety, such as near-misses, poor prognoses, unexpected health outcomes, a patient's deteriorating condition, preventable and unavoidable adverse events, and the unexpected death of a patient, to those more associated with working conditions or structural factors. These include a lack of resources, high care pressure, elevated staff turnover, increased demand, violence against professionals, and moral injury. They can have an emotional impact on professionals, which affects their habits, practices, or clinical decisions [5]. When this behavior change occurs in the undesired direction, we may encounter defensive practices, burnout, or detachment that often entail assuming unnecessary risks for the patient and additional costs for the system [6,7].

Vicious Cycle

Even in the absence of behavioral change, the loss of professional well-being is accompanied by an exponential increase in the risk of human error in patient care, leading to lower performance and a deterioration in the quality of care provided to patients [8]. This worsening of outcomes can, in turn, contribute to the occurrence of new errors and safety incidents. Thus, we find workers in a vicious cycle where the loss of well-being, the inability to regain balance and face stressful situations, and the quality of care are intimately related [5,9]. Hence, preserving and enhancing the resilience of professionals is a key element in ensuring the proper functioning of health care systems, the quality of care, patient safety, and achieving better health outcomes [10].

Resilience

Resilience refers to the ability of individuals, teams, and organizations to adapt, recover, and grow in the face of stress, adversity, and challenging circumstances [11]. In the context of health care, it enables professionals to maintain their well-being and continue to provide high-quality care despite the inevitable challenges they encounter. This not only benefits the individuals themselves but also enhances the safety and effectiveness of care provided by health care teams [12].

Individual Differences

In addition to the inherent risk in health care activities, other cultural aspects can also be a source of distress for health care professionals and may even constitute a barrier to the process of recovering well-being once it has been compromised and are also directly related to values. From an individual perspective, health care professionals share a set of ideals focused on doing good and prioritizing the well-being of others. While these ideals are desirable for good performance, their rigid application in critical situations can intensify distress or hinder constructive coping and professional recovery [13].

Organizational Culture

At the organizational level, the institution's culture plays a key role in determining how professionals feel, interpret, and cope with stressful situations. The prevailing professional culture in the health care sector is characterized by strong ideals of perfection that often lead to a negative and unconstructive view of human response, considering someone weak for asking for help or being unable to cope with a stressful situation. Understanding these causes allows for designing and implementing barriers that prevent or reduce the likelihood of succumbing to stressful situations occurring in the future.

Challenge

The World Health Organization has incorporated resilience in health care systems into the Global Patient Safety Action Plan 2021-2030 through strategy 2.4: providing a strong human factor and ergonomics-based perspective and contributing to strengthening the resilience of health care organizations and clinical practices [14]. This initiative aims to address a basic need for the proper functioning of health care systems, which is to take care of those who care.

We intend to address this need not only from a reactive approach when things go wrong but also by creating resilient teams and providing middle managers with tools to reinforce the resilience of professionals, teams, and the health care organization to ensure their ability to respond to critical situations. We believe that middle managers, due to their strategic position in the institution's structure, are key agents for the development of resilience at all levels; therefore, the materials of this project are specifically aimed at this group.

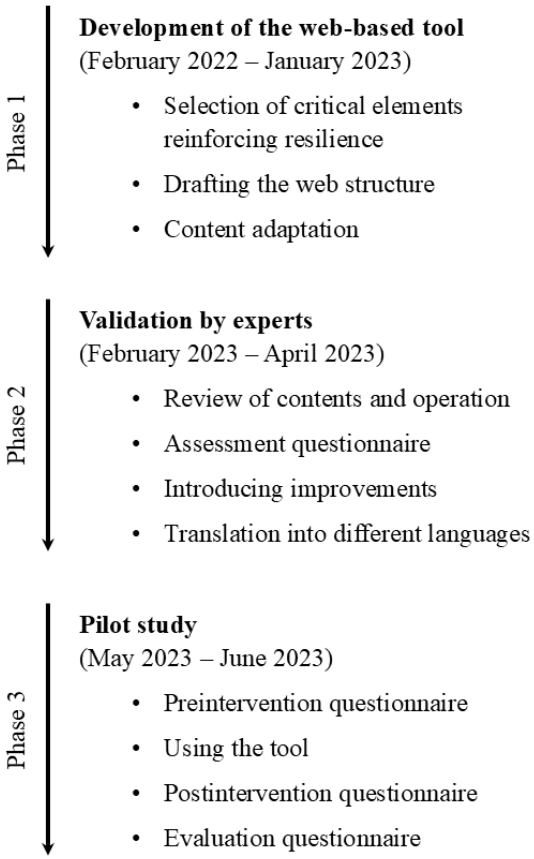
Methods

Study Design

An observational study was developed in 3 phases: development of the web-based tool (phase 1), validation by experts (phase 2), and piloting with professionals (phase 3; Figure 1). Primary care and hospital were defined as usual settings for the new tool

named eResiliencia. The established 8 principles for learning tools that aim to support the translation of resilience into practice by Haraldseid-Driftland et al [15] were used. In this study, resilience in health care was defined as the capacity to adapt to challenges and changes at different system levels to maintain high-quality care [12].

Figure 1. Flowchart of project phases and steps.



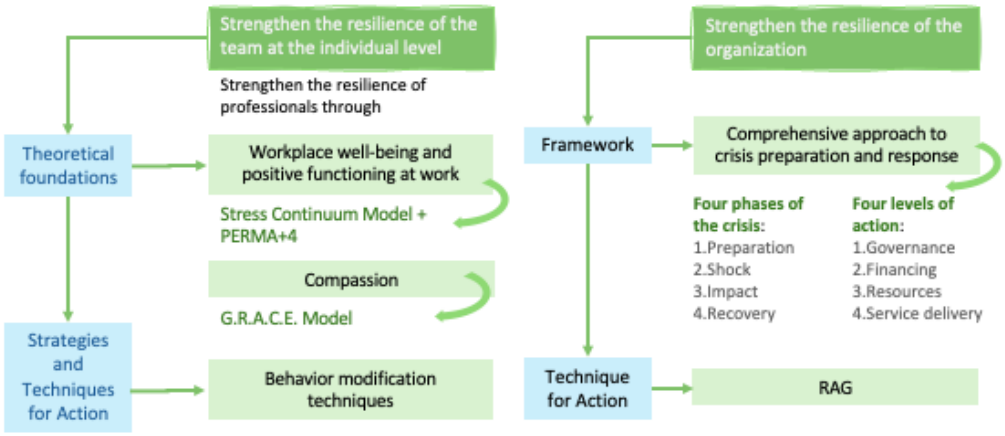
The study was developed by a core group involving a multidisciplinary team supported by a panel of international experts in health administration, occupational safety, and patient safety. It was initiated in February 2022 and concluded in June 2023.

Phase 1: Development of the Web-Based Tool

To design the content for the web-based tool (phase 1), a comprehensive search of critical elements reinforcing resilience was conducted, culminating in a preliminary selection from among the most frequently cited elements. This included the revision of reviews and systematic reviews (after 2019) on the concept of resilience, experimental studies measuring the effectiveness of recent programs designed to enhance the resilience of health care professionals, and the content of different scales for measuring resilience. In particular, this search was aimed at identifying experiences and successful strategies [16].

The eResiliencia intervention is built on a robust theoretical framework that integrates principles of resilience, human behavior, and organizational adaptability to address the complex challenges faced by health care systems (Figure 2). This approach emphasizes strengthening resilience through the development of 5 essential traits: self-control, which involves regulating emotions and behavior under pressure; adaptability or the flexibility to adjust to changes and demands; optimism, characterized by a positive outlook and confidence in overcoming challenges; self-sufficiency, reflecting the ability to independently manage difficulties; and persistence, the sustained effort to persevere in the face of adversity.

Figure 2. eResiliencia theoretical framework. PERMA+4: PERMA+4 Well-Being model; RAG: Resilience Analysis Grid.



The intervention draws from the 5-tier model proposed by Seys et al [17], which combines preventive and supportive measures to build resilience. Levels 1 and 2 focus on proactive strategies such as prevention and self-care, while levels 3 to 5 offer structured peer and clinical support to address resilience at both individual and organizational levels. This multilevel approach ensures that resilience-building efforts are comprehensive and adaptable to the diverse contexts within health care settings.

eResiliencia operates across 3 interconnected levels: individual, team, and organizational. It uses both proactive and reactive strategies to enhance resilience. Proactive measures aim to build capacity before crises occur, using tools such as stress management models and workplace well-being frameworks. Reactive strategies focus on recovery and learning following critical events, ensuring that individuals and organizations can adapt effectively to challenges.

The intervention is informed by several evidence-based models, including the stress continuum model [18], which helps monitor and manage stress; the PERMA+4 Well-Being model [19], which promotes workplace well-being through its multidimensional framework; the G.R.A.C.E. model [20], which supports health care workers in cultivating compassion and managing burnout; and the Resilience Analysis Grid [21], which was applied to assess and enhance team response capacity in unexpected events. In addition, eResiliencia incorporates a set of behavior change techniques [22] designed to equip middle managers with targeted strategies for promoting workplace well-being. These techniques are tailored to the unique cultural and organizational contexts of health care systems, ensuring their relevance and practicality.

Following this, a second step involved drafting the web structure to accommodate the intended contents effectively. Subsequently, a process of content adaptation to the health care context ensued, wherein various support materials for the tool were developed. These materials underwent 2 rounds of internal review before arriving at version 1.0, ensuring their alignment with the intended objectives. The first version of the web platform was developed in Spanish and hosted on the Miguel Hernández University of Elche (Spain) servers, which comply with security requirements and regulations. A total of 11 months were spent to complete this phase, commencing in February 2022.

Phase 2: Validation by Experts

Version 1.0 underwent a review and validation process by an international panel of experts between February and April 2023. A total of 20 professionals integrated this panel, comprising 14 (70%) from Spain, 2 (10%) from Colombia, 1 (5%) from Argentina, 1 (5%) from Ecuador, 1 (5%) from Brazil, and 1 (5%) from Chile. The experts' identification was made using the snowball approach. All experts were informed of the pursued objective and the approximate time commitment and were requested voluntary, unpaid participation. Of the 20 professionals, 11 (55%) were women and 9 (45%) were men, and all met the criterion of a minimum of 10 years of professional experience. It was ensured that no personal evaluation would be disclosed, and communications with them were conducted via telephone or email.

Following a comprehensive examination of all contents and their structural organization, experts engaged were required to respond to a rubric designed to evaluate various aspects. This rubric used a scale ranging from 1 to 5 to assess the clarity of the content and its utility and applicability, grading from completely disagree to completely agree on whether the materials facilitated learning strategies to foster the resilience of professionals or teams under their purview, their likelihood to recommend the tool to other colleagues, and their intention to apply learned tools or strategies. In addition, experts were encouraged to identify strengths and areas for improvement and to rate their overall satisfaction with the web-based intervention and its contents on a scale from 1 to 10, where the higher the score, the greater the satisfaction. The level of agreement among evaluators was assessed to inform decisions regarding aspects requiring changes. All feedback and suggested improvements were meticulously implemented to enhance the quality of the content. Subsequently, the refined version of the tool was translated into English, Portuguese, and Valencian. Careful attention was given to ensuring equivalence across all versions through a process of back translation. Necessary adjustments were made to ensure the adequacy and coherence of the system across languages. An adaptation was also made to the idioms

and expressions commonly used in Latin American Spanish. This procedure also ensured that social and cultural aspects specific to each country were considered.

Phase 3: Pilot Study

From May 1 to June 30, 2023, a pilot study was conducted in a real setting in Ecuador. The participants' feedback served to refine the platform and the tool. A broader cohort of 458 health care professionals from Ecuador was invited to participate. They served as executives or managers in one of the country's public health care networks. This invitation was extended through the Institutional Course on Continuous Improvement in Health Care Quality and Safety offered by the Ministry of Public Health of Ecuador with the support of the Miguel Hernández University, Spain. The invitation was formally extended to all professionals from the management teams of public health care institutions in the country, including physicians, nurses, and other professional profiles. They assessed whether they wished to participate in this training, ensuring their data and performance were completely anonymized. Participants were required to complete a preintervention questionnaire at the moment of registration to access the contents of the web-based tool. This questionnaire comprised 20 questions assessing the knowledge areas to be addressed. In Multimedia Appendix 1, the topics addressed by each question can be consulted. After 7 days, a postintervention questionnaire containing the same questions was made available to assess content assimilation. Participants had the option to retake this postintervention questionnaire up to a maximum of 10 times, with a minimum score of 8 out of 10 required to qualify for a proficiency certificate. Following the completion of the postintervention questionnaire, participants were presented with an evaluation questionnaire consisting of the same questions posed to the experts during phase 2.

Data Analysis

In this study, both quantitative and qualitative analyses were conducted. The profile of participants who completed all phases of the iteration was compared with those who dropped out and did not finish to identify potential differences. For phase 2, descriptive analyses were performed on the rubric scores, and content analyses were conducted for open-ended responses. Fleiss κ was used to analyze the level of agreement among evaluators. The response options were recoded into 2 categories: "agree" (values 4 and 5) and "disagree" (values 1, 2, and 3), evaluating the overall agreement among experts across the 5 variables (clarity, utility, intention to recommend, intention to apply, and eResiliencia).

Similarly, in phase 3, the same analyses as in phase 2 were performed for the evaluation questionnaire. Descriptive statistics were conducted for both the preintervention and postintervention questionnaires. To assess differences between the means of such questionnaires, the Wilcoxon signed rank test was used. In addition, the McNemar statistic was calculated to compare the number of correct responses before and after the use of the web-based tool for each question. Qualitative analysis also considered the frequency of positive comments and improvement suggestions made by participants at the end of the structured survey. These comments, expressed in natural language, were categorized by 2 researchers (EG-H and IC) to facilitate their documentation. Finally, sociodemographic data were analyzed using descriptive statistics.

Ethical Considerations

In accordance with the Ecuadorian Ministerial Agreement AM00005-2022 and the Spanish Law on Biomedical Research (14/2007), studies conducted through anonymous surveys that do not collect health-related data are exempt from ethical review. This study was fully compliant with all regulatory standards for personal data protection. Ethical principles for medical research involving human participants included in the guidelines set by the Helsinki Declaration were followed. Informed consent for study participation was obtained at the time of registration on the platform, whereby individuals were required to select the corresponding checkbox, with instructions provided regarding the process for revoking their participation. No form of financial compensation was provided for participation.

Results

Phase 1: Development of the Web-Based Tool

A total of 42 documents were deemed relevant and were analyzed in depth to extract lessons learned and proposals for the design of eResiliencia (Table 1).

Table 1. Documents considered when designing the content of eResiliencia.

Name	Study
Resilience scales (n=7)	
Resilience at Work Scale	Winwood et al [23], 2013
50-Item Resilience Questionnaire	The Psychometric Project [24], 2013
Brief Resilience Scale	Smith et al [25], 2008
Connor-Davidson Resilience Scale	Connor and Davidson [26], 2003
Resilience Scale for Adults	Friborg et al [27], 2003
Resilience Scale	Wagnild and Young [28], 1993
Dispositional Resilience Scale	Bartone et al [29], 1989
Resilience concept (n=11)	
Psychological Resilience: An Affect-Regulation Framework	Troy et al [30], 2023
A Simultaneous Concept Analysis of Resilience, Coping, Posttraumatic Growth, and Thriving	Bowling et al [31], 2022
An Approach to the Unified Conceptualization, Definition, and Characterization of Social Resilience	Moya and Goenechea [32], 2022
Integrative Review of the Recent Literature on Human Resilience: From Concepts, Theories, and Discussions Towards a Complex Understanding	Métais et al [33], 2022
Resilience: An Integrated Review	Daly [34], 2020
Psychological Resilience as an Emergent Characteristic for Well-Being: A Pragmatic View	Tay and Lim [35], 2020
What Does Resilience Signify? An Evaluation of Concepts and Directions for Future Research	Infurna [36], 2020
Nurse Resilience: A Concept Analysis	Cooper et al [37], 2020
What Are the Factors Affecting Resilience in Health Professionals? A Synthesis of Systematic Reviews	Huey and Palaganas [38], 2020
Toward a Transversal Definition of Psychological Resilience: A Literature Review	Sisto et al [11], 2019
Resilience as a Multimodal Dynamic Process	Stainton et al [39], 2019
Resilience in health care systems (n=9)	
Exploring the Nature of Adaptive Capacity for Resilience in Healthcare Across Different Healthcare Contexts; a Metasynthesis of Narratives	Lyng et al [40], 2022
Resilience in Organization-Related Research: An Integrative Conceptual Review Across Disciplines and Levels of Analysis	Raetze et al [41], 2022
Shifting Focus from Burnout and Wellness toward Individual and Organizational Resilience	Vercio et al [42], 2021
Health System Resilience: a Literature Review of Empirical Research	Biddle et al [43], 2020
Measuring the Resilience of Health Systems in Low- and Middle-Income Countries: a Focus on Community Resilience	Bhandari and Alonge [44], 2020
Maintaining capacity in the health care system during the COVID‐19 pandemic by reinforcing clinicians' resilience and supporting second victims	Strametz et al [45], 2020
COVID-19: Peer Support and Crisis Communication Strategies to Promote Institutional Resilience	Wu et al [46], 2020
Strengthening Health Systems Resilience: Key Concepts and Strategies	Thomas et al [47], 2020
Conceptual Analysis of Health Systems Resilience: A Scoping Review	Turenne et al [48], 2019
Existing intervention programs to enhance resilience (n=6)	
Psychosocial Interventions for Building Resilience of Informal Carers of People Living with Stroke: a Systematic Review	Qureshi et al [49], 2023
Building Personal Resilience following an online Resilience Training Program for BScN Students	Stoliker et al [50], 2022

Name	Study
A Resilience-Building App to Support the Mental Health of Health Care Workers in the COVID-19 Era: Design Process, Distribution, and Evaluation	Golden et al [51], 2021
The Effectiveness of Charge Nurse Training on Leadership Style and Resiliency	Spiva et al [52], 2020
Comprehensive Meta-analysis of Resilience Interventions	Liu et al [53], 2020
Interventions to Improve Resilience in Physicians Who have Completed Training: A Systematic Review	Venegas et al [54], 2019
Useful tools (n=9)	
PERMA+4: A Framework for Work-Related Wellbeing, Performance and Positive Organizational Psychology 2.0	Donaldson et al [19], 2022
Effectiveness of a One Day Self-Compassion Training for Pediatric Nurses' Resilience	Franco and Christie [55], 2021
The Positive Functioning at Work Scale: Psychometric Assessment, Validation, and Measurement Invariance	Donaldson and Donaldson [56], 2021
The PERMA-Profiler: A Brief Multidimensional Measure of Flourishing	Butler and Kern [57], 2016
The Behavior Change Technique Taxonomy (v1) of 93 Hierarchically Clustered Techniques: Building an International Consensus for the Reporting of Behavior Change Interventions	Michie et al [22], 2013
G.R.A.C.E. for Nurses: Cultivating Compassion in Nurse/Patient Interactions	Halifax [20], 2014
RAG—Resilience Analysis Grid	Hollnagel et al [21], 2011
US Marine Corps and Navy Combat and Operational Stress Continuum Model: A Tool for Leaders	Nash [18], 2016
Caring for Our Own: Deploying a Systemwide Second Victim Rapid Response Team	Scott et al [58], 2010

The working team outlined a presentation structure for eResiliencia and chose the most suitable appearance from several options. With the information gathered and the ideas from the working group itself, the first draft of the content was structured in the beta version of the website that hosts eResiliencia [59].

The content was divided into 2 main blocks (Figure 3), one aimed at building resilience at the individual and team levels, and another at the organizational level, preceded by a conceptual framework. Platform navigation was conducted through both the menu bar and the side panel (Figure 4). Within each block, there were subdivisions as shown in Table 2. In the tab dedicated to the conceptual framework, an explanation of the course context, an introduction to resilience, several examples of situations where it is crucial, and a section with definitions of the most relevant concepts of the course were provided. The first block, tools to strengthen resilience at the individual and team levels, offered the necessary strategies to enhance the resilience of team members at an individual level, placing particular emphasis on the importance of workplace well-being. Finally, the block on tools to reinforce resilience at the organizational level provided strategies to assess and achieve resilient institutions in the face of crises and unexpected events. In addition, a road map was included at the end of the course to aid in selecting the most useful strategies based on the objectives.

Figure 3. The home screen of the eResiliencia web-based platform.

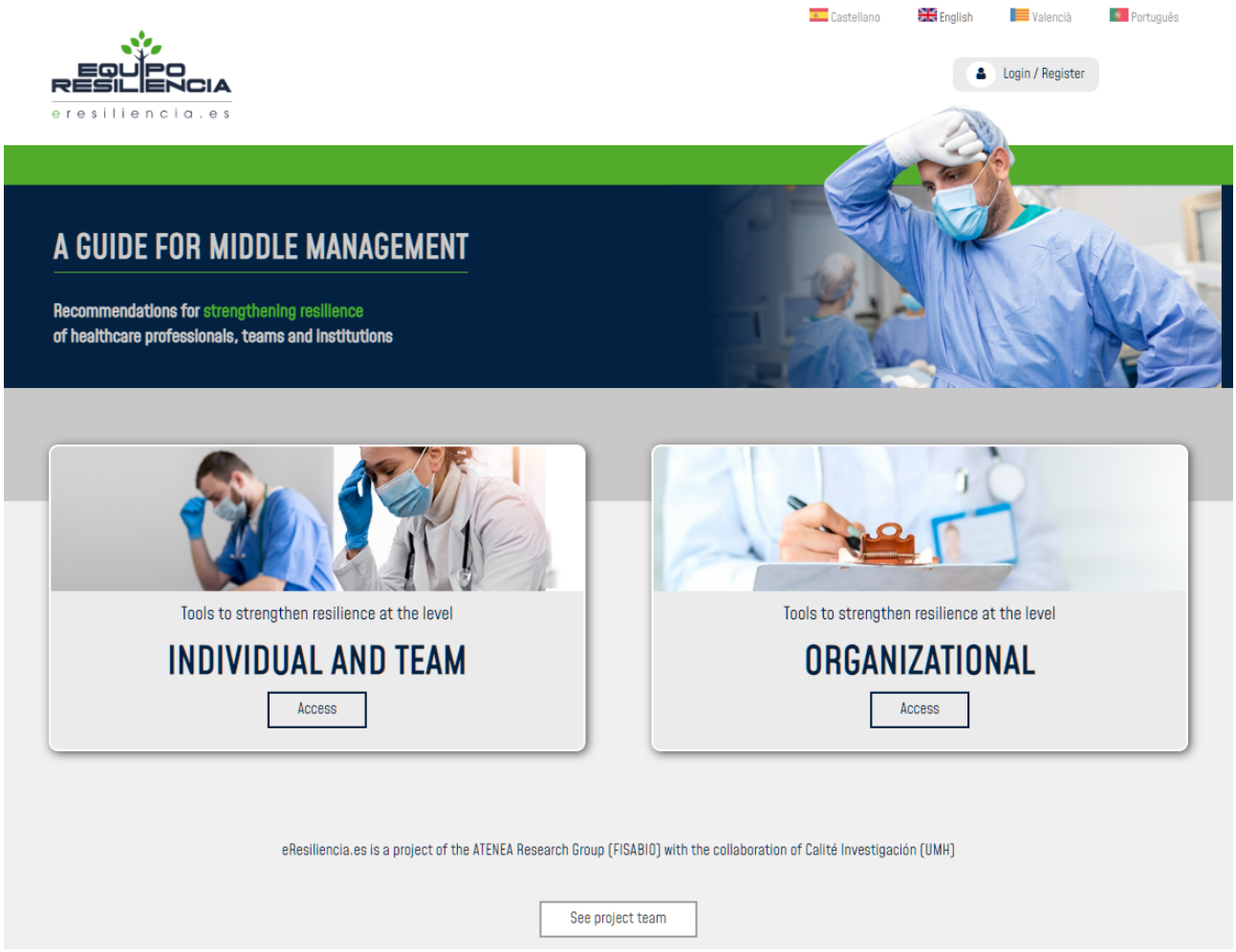


Figure 4. Navigation panel of the eResiliencia web-based platform.

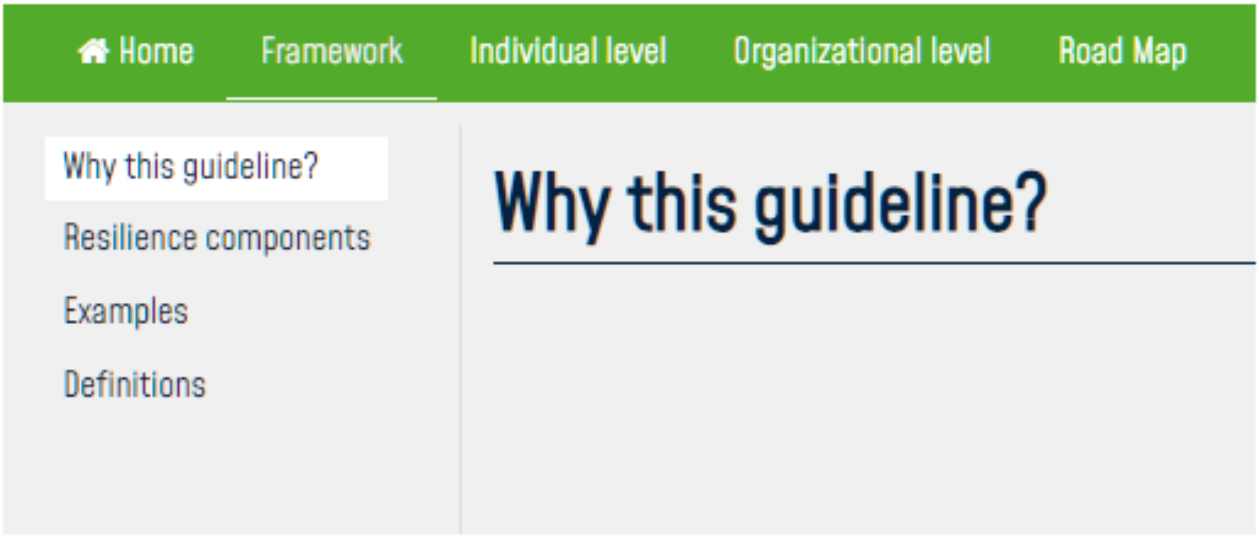


Table 2. Structure of the eResiliencia web-based platform.

Structure	Items
Framework	• Why this guideline? • Resilience components • Examples • Definitions
Individual level	• Stress Continuum Model • Enhancing resilience at the individual level • G.R.A.C.E. model • PERMA+4 Well-Being model for occupational wellness • Behavior change technique for building work-related well-being
Organizational level	• Enhancing resilience at the organizational level • Resilience Analysis Grid

In total, it comprised 17 videos, 11 figures and images, 4 PDF files, and references for further information expansion, among which downloadable materials or video, graph, and text links were also available. Upon accessing their personal account, users were provided with a section containing pertinent information regarding the platform's functionality, along with a video elucidating navigation and content. In addition, users had access to a support resource and the option to change their password.

Phase 2: Validation by Experts

Of the 20 experts, 13 (65%) assessed the tool and provided feedback. They rated the clarity of the content at an average of 4.5 (SD 0.7) and the utility and applicability at an average of 4.7 (SD 0.5), out of 5 in both cases. Regarding whether the materials facilitated learning strategies to foster the resilience of professionals or teams under their purview, 62% (8/13) of the experts indicated they *completely agree*, 31% (4/13) of the experts indicated they *agree*, and 8% (1/13) of the experts indicated they *completely disagree*. Regarding their likelihood to recommend the tool to other colleagues, 92% (12/13) of the experts responded *definitely yes*, and 8% (1/13) of the experts responded *probably yes*. The same results were obtained for their intention to apply learned tools or strategies. The level of agreement among experts for the combined set of the 5 evaluative variables of eResiliencia was 0.79 (95% CI 0.65-0.92).

Overall satisfaction was rated with an average of 9.3 (SD 0.6) out of 10 points. Some of the strengths highlighted were the chosen content and the ease of navigation and presentation of the web-based platform. In contrast, areas for improvement included making the videos more engaging and the possibility of incorporating practical exercises or a forum for participant interaction. As a suggestion, an introductory video of the content and the functionality of the web-based platform was included (Multimedia Appendix 2) [60]. The full list of comments made by experts is presented in Multimedia Appendix 3.

Phase 3: Pilot Study

Of the 458 professionals invited to participate, 362 (79%) registered on the platform. The preintervention and postintervention questionnaires of the pilot study were completed by 218 (60.2%) of these 362 health care professionals exercising leadership roles who began the intervention. Moreover, 146 (40.3%) health care professionals responded to the experience evaluation questionnaire.

Most of them were physicians (90/218, 41.3%) and nurses (59/218, 27.1%), with a mean experience of 10.3 (SD 8.8) years. Their primary work setting was primary care (113/218, 51.8%), hospital (79/218, 36.2%), quality area (9/218, 4.1%), or other management positions (7/218, 3.2%). The profile of those who did not complete the entire educational intervention was similar to those who did, with 42.4% (61/144 being physicians, 24.3% (35/144) being nurses, and 45.1% (65/144) working in primary care. Those who did not complete the intervention had slightly average experience, with 11.4 years compared with 10 years.

The scores obtained in the preintervention questionnaire averaged 42.0 (SD 16.6) of 100, while for the postintervention questionnaire, they averaged 74.7 (SD 22.7), with significant differences between them (P<.001; Z=–11.728). The increase in the percentage of correct answers for each question is represented in Table 3.

Table 3. Increase in percentage of correct answers for each question (n=218).

Questions[a]	Correct responses on the preintervention questionnaire, n (%)	Correct responses on the postintervention questionnaire, n (%)	Increase of correct answers (%)	McNemar test	*P* value
1	148 (67.9)	192 (88.1)	20.2	25.681	<.001
2	73 (33.5)	148 (67.9)	34.4	53.165	<.001
3	102 (46.8)	155 (71.1)	24.3	27.876	<.001
4	32 (14.7)	134 (61.5)	46.8	87.940	<.001
5	103 (47.2)	165 (75.7)	28.4	37.969	<.001
6	66 (30.3)	176 (80.7)	50.5	99.008	<.001
7	92 (42.2)	162 (74.3)	32.1	52.900	<.001
8	110 (50.5)	180 (82.6)	32.1	48.582	<.001
9	94 (43.1)	153 (70.2)	27.1	30.862	<.001
10	82 (37.6)	159 (72.9)	35.3	55.010	<.001
11	139 (63.8)	189 (86.7)	22.9	29.280	<.001
12	117 (53.7)	169 (77.5)	23.8	28.900	<.001
13	44 (20.2)	165 (75.7)	55.5	111.628	<.001
14	110 (50.5)	181 (83)	32.5	49.495	<.001
15	29 (13.3)	111 (50.9)	37.6	64.324	<.001
16	86 (39.4)	145 (66.5)	27.1	33.307	<.001
17	138 (63.3)	194 (89)	25.7	34.375	<.001
18	114 (52.3)	154 (70.6)	18.3	20.554	<.001
19	84 (38.5)	162 (74.3)	35.8	55.934	<.001
20	68 (31.2)	163 (74.8)	43.6	73.025	<.001

[a]Because this is active training, the questions have been masked to avoid spoilers. They are available upon reasonable request.

Moreover, 58.3% (127/218) of the health care professionals passed the postintervention questionnaire without requiring any retries. However, 14.2% (31/218) of the health care professionals needed 1 retry, 8.3% (18/218) needed 2 retries, 6% (13/218) needed 3 retries, and 3.7% (8/218) needed between 4 and 7 retries. Furthermore, 9.6% (21/218) of the health care professionals exhausted all 10 retry attempts without achieving the required score.

Regarding the perception of the 146 people who completed the evaluation questionnaire, the clarity of the content received an average score of 4.7 (SD 0.6) on a 1 to 5 scale, while the utility or applicability scored 4.8 (SD 0.5). The results of their opinion are depicted in Table 4.

Global satisfaction was rated 9.4 (SD 1.1) out of 10 points on average. Among the strengths underscored were the comprehensiveness of the content and the interesting techniques explained. Conversely, identified improvement areas consisted of incorporating more illustrative examples and providing downloadable materials.

Table 4. Results of the nonnumerical scale questions from the assessment questionnaire (n=146).

Question	Completely disagree, n (%)	Disagree, n (%)	Neither agree nor disagree, n (%)	Agree, n (%)	Completely agree, n (%)	Definitely not, n (%)	Probably not, n (%)	Not sure, n (%)	Probably yes, n (%)	Definitely yes, n (%)
The course content has enabled me to learn strategies for reinforcing the resilience of the professionals or teams under my supervision.	7 (4.8)	1 (0.7)	6 (4.1)	54 (37)	78 (53.4)	—[a]	—	—	—	—
I will recommend the course to other colleagues.	—	—	—	—	—	0 (0)	0 (0)	6 (4.1)	22 (15.1)	118 (80.8)
I will apply some of the tools or strategies learned to strengthen individual, team, or organizational resilience in my work environment.	—	—	—	—	—	0 (0)	2 (1.4)	2 (1.4)	28 (19.2)	114 (78.1)

[a]Not applicable.

Qualitative Analysis of Open-Ended Comments

Strengths

The most repeated topics among the strengths were the learning of new techniques or specific models (34/146, 23.3%), the web-based methodology that allowed access at any time (14/146, 9.6%), the organizational approach (10/146, 6.8%), the clarity of the content (9/146, 6.1%), stress management (6/146, 4.1%), and the examples provided (4/146, 2.7%).

Opportunities for Improvement

In terms of aspects to be improved, the incorporation of more examples (14/146, 9.6%) and the expansion of the content (14/146, 9.6%) were generally highlighted. In addition, improving the quality of the videos (12/146, 8.2%), including the possibility of interacting with other students in a more practical way (9/146, 6.2%), enabling downloadable content (9/146, 6.2%), reducing the length of the questionnaires (6/146, 4.1%), including web-based tutorials for resolving doubts with the teachers (5/146, 3.4%), and adding a progress bar to visualize the course as a whole (4/146, 2.7%) were emphasized.

Discussion

Principal Findings

Our developed web-based approach operates as a comprehensive tool designed primarily for middle managers, aiming to foster resilience at both individual and organizational levels within the staff at their charge. By using multimedia resources, this approach has been shown to enhance the knowledge about resilience of health care professionals across diverse fields and backgrounds.

eResiliencia was highly rated by the panel of experts regarding clarity and usefulness, indicating that the tool is perceived as comprehensible and relevant for application in professional contexts. High ratings were also obtained for the willingness to recommend the tool to colleagues and the intention to apply the learned strategies, demonstrating strong acceptance and confidence in the resource. For most (12/13, 92%) participants, the materials were potentially effective in strengthening team resilience, except for 1 (8%) expert who expressed complete disagreement. This highlights the need to consider diverse perspectives and possible improvements in adapting the content to various contexts. Nonetheless, overall satisfaction was high, with more than 9 points out of 10.

Implementing eResiliencia in Real-World Settings

The eResiliencia platform features a user-friendly interface, and in most cases, managerial staff possesses the necessary skills to use such digital tools effectively. However, some systemic limitations may impact its implementation. First, the increasing measures to prevent cyberattacks and data ransom incidents in health care institutions have led to stricter internet access controls, even in areas where connectivity was previously unrestricted. This decision may hinder access to eResiliencia within some organizations. In addition, in certain health care networks, managers lack the resources to access the platform outside their offices. Incorporating offline functionality could address these limitations and facilitate broader access to the platform.

Second, resistance from health care organizations, often rooted in established workflows or skepticism about the benefits of new interventions, represents a traditional barrier. In this case, the integration of eResiliencia was facilitated by embedding it within a national patient safety and quality training program aimed at launching a global institutional strengthening plan. A similar strategy could help overcome organizational resistance in other settings.

Third, ensuring flexibility and accessibility is essential to meet the diverse needs of users. The use of asynchronous learning options was key in this instance, allowing participants to decide when and how to engage with the platform. This flexibility significantly increases accessibility and is particularly relevant in environments with high workloads. However, allocating dedicated time during work hours could further enhance participation rates by reducing competing demands on participants' schedules.

Comparison With Prior Studies

It is known that health care professionals face stressful situations in their professional work. The frequency of burnout or the existence of defensive practices is also not uncommon. It is no surprise that the role of team leaders influences the team's ability to respond and recover after experiencing adverse situations [61]. The COVID-19 pandemic brought to the attention of managers, executives, and society in general concerns about the resilience capacity of these collectives, both at an individual and collective level, to face what has undoubtedly been the greatest global challenge for health care professionals. This study originates from that period and delves into both aspects, offering an easily accessible training format for those responsible for clinical teams. In this regard, it is based on the observation that most programs designed to increase resilience were developed to operate at the individual level and that it was necessary to create interventions focused on the work group [30].

eResiliencia incorporates agreed-upon resilience attributes [37] and skills that leaders must possess to strengthen their teams' resilience, which include self-efficacy, self-control, the ability to provide support and help, and learning from difficulties [62].

Uncertainty, resource shortages, constant protocol variations, the risk of infection and transmission, and critical decision-making with ethical implications, among other factors, were prevalent during the COVID-19 pandemic and contributed to fear, anxiety, stress, moral injury, or compassion fatigue among health care professionals [63,64]. These situations have highlighted the need to reconsider the role of middle management, as they must address the needs of their teams to face challenges. Therefore, their resilience is a crucial competence required for their work.

The scores obtained in this study by the professionals who underwent the training show that after the use of the tool, the results were better both overall and for each individual question. However, the preintervention questionnaire also revealed that participants' previous knowledge was relatively low in many areas, with correct response rates ranging from 13.3% (question 15; behavior change technique) to 67.9% (question 1; definition of resilience). This underscores the necessity and effectiveness of the educational intervention implemented. The greatest increases were in questions with initially low correct response rates (question 4, question 6, and question 13), suggesting the acquisition of new concepts. In contrast, questions such as question 1 and question 18, which initially had a higher percentage of correct responses, showed a less pronounced increase compared to those with lower initial performance. The increase in scores on the final questionnaire compared to the initial questionnaire is consistent with the findings of other studies, which show that both in-person [52,65-67] and on the web [50,68] interventions increase resilience capacity. In addition, it has been observed that they also protect the mental health of professionals [69].

Solutions to assist professionals were implemented in different countries and were similar, including the availability of psychological services, support hotlines, and websites with resource materials [70]. Focusing on tools available for Spanish speakers, an example is the digital platform and mobile app, Be+Against COVID, which was used in Argentina, Brazil, Colombia, Chile, Ecuador, and Spain [71].

Although middle managers have usually received training in occupational risk prevention and conflict management, many times there is a lack of specific and detailed training in fundamental aspects of team management and resilience. Traditionally, team management has focused on organizational and clinical aspects to ensure that patients' needs are adequately met. Nowadays, middle managers are expected to efficiently manage the resources of the professional teams they lead, particularly human talent, to adequately respond to health challenges.

Implications of Findings

In the postpandemic era, there is a growing consensus that among the functions of executives and middle managers (such as the organization of responses to care demands, task allocation, conflict management within teams, and the strengthening of teams to meet care demands effectively), there should be an emphasis on maintaining and strengthening the resilience capacity of their teams of professionals, because their ability to cope with highly stressful situations determines the quality of care provided and the outcomes for patients. This eResiliencia program follows this trend and takes a proactive approach by offering a framework for executives and middle managers to use some or all the resources to achieve this objective and prepare their team for everyday challenges and future crises.

Although most (197/218, 90.4%) of the participants were able to pass the training without any difficulties, some (21/218, 9.6%) individuals exhausted all reattempts without achieving the minimum required score. While these data suggest that the material and teaching methods are effective for most, it would be beneficial to delve into the reasons behind the difficulties faced by this small percentage to address and mitigate the problem.

Resilience training is presented as a crucial element in the professional development of middle managers in the health care sector. Promoting and reinforcing resilience is not only essential for the individual well-being of professionals but also strengthens the cohesion and effectiveness of the team as a whole. However, for this to be feasible, it is essential that clinical management evolves and provides effective tools to middle managers, facilitating resource management, including resilience training.

In this era, there is a growing consensus that the existing "find and fix" solutions should be replaced by those that are more forward-thinking [15]. Moreover, the organizational culture is

critical to ensuring patients receive adequate care and workers' resilience when things go wrong [72]. Health care organizations are evolving toward the principles of Safety II framework here [73], which involves recognizing the challenge of complexity and uncertainty in clinical practice to anticipate errors. Promoting the resilience of workers at both individual and team levels contributes to realizing the implementation of Safety II [74].

Fragmentation in the assignment of professionals poses a significant challenge to teamwork. Thus, cohesion among middle medical and nursing managers is essential to overcoming the barriers created by this fragmentation and fostering a more collaborative and efficient work environment. The training of middle managers in the health care organization must be a top-down strategy, driven by leadership. Health care executives must communicate these priorities to middle managers and provide the necessary support for implementation. Only through strong and consistent leadership from the upper echelons can we ensure that middle managers are equipped with the competencies required to effectively manage their teams and promote a resilient work environment.

The response rate indicates that 4 (40%) of the 10 professionals who registered on the platform were eventually not interested in the training on how to foster resilience in their teams and collaborators. This could be because they already possess the personal resources to tackle the task or because eResiliencia did not meet their expectations or failed to motivate them sufficiently. This aspect could be addressed, for example, with a more practical approach through group dynamics and cooperative tasks to generate greater interest and engagement, so participants can see a more direct application of the training tools.

Strengths

This study addresses the 5 key questions of a resilience study in the health care sector [75]. It proposes an alternative approach to individual-focused training by emphasizing the role of clinical team leaders and how to strengthen the resilience of their teams and team members.

Limitations

Although international experts share a common cultural background, they represent a group of Spanish-speaking countries. The generalization of content to other countries should not be done directly. Participants in the pilot study phase worked in various public health networks in Ecuador and took part in this experience as part of a quality and patient safety management course, which emphasized the need to move away from a blame culture. This aspect may have influenced their engagement with the eResiliencia content. Participants had a set amount of time to complete the training on the eResiliencia web-based platform. In a different self-directed learning context without this limitation, trainee behavior might be different. Because this is a study without a control group, the changes observed in participants' knowledge may not be exclusively attributable to eResiliencia. Finally, it should be noted that although the participants were enrolled in training organized by the Ministry of Public Health of Ecuador, it was emphasized that they could freely choose whether to complete the eResiliencia training and respond to the satisfaction survey. Answering the questionnaire did not provide any advantage and the responses were not disclosed, ensuring anonymity at all times, which was clearly communicated to the participants. This approach was taken to avoid a reporting bias.

Cultural Factors Influencing Resilience-Building Capacity

Ecuador's health care system is characterized by hierarchical structures and centralized decision-making, typical of high-power distance cultures. eResiliencia capitalizes on the influential role of middle managers, who act as intermediaries between upper management and frontline workers. By equipping middle managers with resilience-building tools, the intervention empowers them to cascade these strategies throughout their teams, reinforcing resilience at all levels of the organization. The networks engaged in this study operate with limited resources, requiring flexibility and improvisation to address challenges. eResiliencia provides structured yet adaptable strategies, offering tools for middle managers to tailor resilience-building efforts according to the specific needs and constraints of their teams. Although Ecuador faces challenges in long-term health care planning, eResiliencia introduces structured frameworks that emphasize proactive resilience building. These frameworks align with preventive measures to prepare health care professionals for crises, fostering a shift from reactive to proactive management.

Many of these aspects are common to countries in this region. However, the outcomes of eResiliencia in Asia, Europe, or North America are likely to differ significantly due to the cultural factors influencing resilience-building capacity. Cultural dimensions such as collectivism, power distance, and approaches to uncertainty, among others, play a crucial role in shaping how resilience strategies are developed and applied across different contexts.

Future Research

It is necessary to delve into a common definition of what is meant by resilience [11,75] and, by extension, the role of managers in the resilience of their team members. A critical aspect for future development involves the incorporation of mechanisms to monitor the utility and long-term impact of the knowledge acquired through the web-based tool, particularly considering the less optimistic data provided by some studies on nursing students [76]. In later phases, this intervention will be complemented by the development of improvement and action plans based on the application of the eResiliencia tools by health care teams and middle management interested in their centers. This will allow a more effective monitoring of the tool, its usefulness, and applicability. These approaches could also be extended to companies that provide home care, to train those who coordinate the work of informal caregivers in the home [77], in light of the growing caregiving economy. Finally, as these are preliminary data, this educational intervention should be tested in the context of other organizational models and diverse cultural settings to validate its effectiveness, acceptability, and utility.

Conclusions

The proposal to evaluate each professional, identify signs of stress, take appropriate measures, and manage team cohesion represents a training initiative that could be implemented clearly and explicitly. However, it is essential to first analyze the organizational context where the intervention will be applied and identify potential adjustments to improve its acceptability based on national and local organizational factors. This tailored approach ensures that the intervention aligns with the specific needs and characteristics of each setting, enhancing its effectiveness and relevance. This training would enable middle managers not only to manage conflicts but also to prevent them and foster a healthy and productive work environment. The developed tool proved to be well received by both the experts and the health care professionals who participated in this pilot phase.

Acknowledgments

The authors would like to acknowledge the experts who participated in the evaluation and validation of the web-based platform as well as all the participants from Ecuador who took part in the pilot study, thereby contributing to an improved experience with the platform. The authors also extend their gratitude to the National Directorate of Quality, Patient Safety, and Health Control of the Ministry of Public Health of Ecuador for their support and dissemination efforts in the initial recruitment process. Clara Pérez contributed to the statistical analyses.

The project was funded by Conselleria de Sanitat Universal i Salut Pública (Generalitat Valenciana, Spain) and the EU Operational Program of the European Regional Development Fund for the Valencian Community 2014-2020, within the framework of the Recovery Assistance for Cohesion and the Territories of Europe program, as the European Union’s response to the COVID-19 pandemic; Ser+ Contra COVID (UGP-21-408). The maintenance of the web-based platform is made possible by the funds allocated to the preparatory action in the UniSalut Program 2023 (subprogram ILISABIO, reference ILISABIO22_AP09). Throughout the execution of this study and composition of this manuscript, JJM benefited from an augmented research activity contract granted by the Carlos III Health Institute (reference INT22/00012).

Data Availability

The datasets generated analyzed during this study are available from the corresponding author on reasonable request. Topics addressed in each question are displayed in Multimedia Appendix 1.

Authors' Contributions

All the authors meet the International Committee of Medical Journal Editors criteria for authorship. JJM was responsible for the design of the study. JJM and IC conducted the search and the preliminary selection of contents related to resilience. IC and EGH drafted the web-based platform and developed the materials. JMD was responsible for the enrollment of people and the acquisition of data. DGT performed the statistical analysis and interpretation of the results. EGH developed the first version of the manuscript. All authors revised the paper critically for important intellectual content and read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topics of the preintervention and postintervention questionnaire questions.
[DOC File , 132 KB - humanfactors_v12i1e67263_app1.doc]

Multimedia Appendix 2
eResiliencia introductory video.
[MP4 File (MP4 Video), 189408 KB - humanfactors_v12i1e67263_app2.mp4]

Multimedia Appendix 3
List of comments received from the experts during the validation phase.
[DOC File , 152 KB - humanfactors_v12i1e67263_app3.doc]

References

1. Cantu J, Gharehyakheh A, Fritts S, Tolk J. Assessing the HRO: tools and techniques to determine the high-reliability state of an organization. Saf Sci 2021 Feb;134:105082. [doi: 10.1016/j.ssci.2020.105082]
2. Schulman PR. Reliability, uncertainty and the management of error: new perspectives in the COVID‐19 era. Conting Crisis Mgmt 2021 Mar 12;30(1):92-101. [doi: 10.1111/1468-5973.12356]

3. Jacobs B, McGovern J, Heinmiller J, Drenkard K. Engaging employees in well-being: moving from the triple aim to the quadruple aim. Nurs Adm Q 2018;42(3):231-245. [doi: 10.1097/NAQ.0000000000000303] [Medline: 29870489]
4. Søvold LE, Naslund JA, Kousoulis AA, Saxena S, Qoronfleh MW, Grobler C, et al. Prioritizing the mental health and well-being of healthcare workers: an urgent global public health priority. Front Public Health 2021 May 07;9:679397 [FREE Full text] [doi: 10.3389/fpubh.2021.679397] [Medline: 34026720]
5. Schiess C, Schwappach D, Schwendimann R, Vanhaecht K, Burgstaller M, Senn B. A transactional "second-victim" model-experiences of affected healthcare professionals in acute-somatic inpatient settings: a qualitative metasynthesis. J Patient Saf 2021 Dec 01;17(8):e1001-e1018. [doi: 10.1097/PTS.0000000000000461] [Medline: 29384831]
6. Zheng J, Lu Y, Li W, Zhu B, Yang F, Shen J. Prevalence and determinants of defensive medicine among physicians: a systematic review and meta-analysis. Int J Qual Health Care 2023 Dec 07;35(4):mzad096. [doi: 10.1093/intqhc/mzad096] [Medline: 38060672]
7. Bourne T, Shah H, Falconieri N, Timmerman D, Lees C, Wright A, et al. Burnout, well-being and defensive medical practice among obstetricians and gynaecologists in the UK: cross-sectional survey study. BMJ Open 2019 Nov 25;9(11):e030968 [FREE Full text] [doi: 10.1136/bmjopen-2019-030968] [Medline: 31767585]
8. Schwappach DL, Boluarte TA. The emotional impact of medical error involvement on physicians: a call for leadership and organisational accountability. Swiss Med Wkly 2009 Jan 10;139(1-2):9-15 [FREE Full text] [doi: 10.4414/smw.2009.12417] [Medline: 18951201]
9. Mira J, Carillo I, Tella S, Vanhaecht K, Panella M, Seys D, et al. The European Researchers' Network Working on Second Victim (ERNST) policy statement on the second victim phenomenon for increasing patient safety. Public Health Rev 2024 Sep 18;45:1607175 [FREE Full text] [doi: 10.3389/phrs.2024.1607175] [Medline: 39360222]
10. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med 2014;12(6):573-576 [FREE Full text] [doi: 10.1370/afm.1713] [Medline: 25384822]
11. Sisto A, Vicinanza F, Campanozzi LL, Ricci G, Tartaglini D, Tambone V. Towards a transversal definition of psychological resilience: a literature review. Medicina (Kaunas) 2019 Nov 16;55(11):745 [FREE Full text] [doi: 10.3390/medicina55110745] [Medline: 31744109]
12. Wiig S, Aase K, Billett S, Canfield C, Røise O, Njå O, et al. Defining the boundaries and operational concepts of resilience in the resilience in healthcare research program. BMC Health Serv Res 2020 Apr 19;20(1):330 [FREE Full text] [doi: 10.1186/s12913-020-05224-3] [Medline: 32306981]
13. Watson P, Westphal R. Stress first aid for health care workers. National Center for PTSD. 2020. URL: https://www.ptsd.va.gov/professional/treat/type/SFA/docs/SFA_HCW_Manual_508.pdf [accessed 2025-01-13]
14. Global patient safety action plan 2021-2030: towards eliminating avoidable harm in health care. World Health Organization. 2021. URL: https://www.google.co.in/books/edition/Global_patient_safety_action_plan_2021_2/csZqEAAAQBAJ?hl=en [accessed 2025-01-13]
15. Haraldseid-Driftland C, Lyng HB, Guise V, Waehle HV, Schibevaag L, Ree E, et al. Learning does not just happen: establishing learning principles for tools to translate resilience into practice, based on a participatory approach. BMC Health Serv Res 2023 Jun 16;23(1):646 [FREE Full text] [doi: 10.1186/s12913-023-09653-8] [Medline: 37328864]
16. Haraldseid-Driftland C, Dombestein H, Le AH, Billett S, Wiig S. Learning tools used to translate resilience in healthcare into practice: a rapid scoping review. BMC Health Serv Res 2023 Aug 23;23(1):890 [FREE Full text] [doi: 10.1186/s12913-023-09922-6] [Medline: 37612671]
17. Seys D, Panella M, Russotto S, Strametz R, Joaquín Mira J, van Wilder A, et al. In search of an international multidimensional action plan for second victim support: a narrative review. BMC Health Serv Res 2023 Jul 31;23(1):816 [FREE Full text] [doi: 10.1186/s12913-023-09637-8] [Medline: 37525127]
18. Nash WP. US Marine Corps and Navy combat and operational stress continuum model: a tool for leaders. In: Combat and Operational Behavioral Health. North Charleston, SC: CreateSpace Independent Publishing Platform; 2016.
19. Donaldson SI, van Zyl LE, Donaldson SI. PERMA+4: a framework for work-related wellbeing, performance and positive organizational psychology 2.0. Front Psychol 2022 Jan 24;12:817244 [FREE Full text] [doi: 10.3389/fpsyg.2021.817244] [Medline: 35140667]
20. Halifax J. G.R.A.C.E. for nurses: cultivating compassion in nurse/patient interactions. J Nurs Educ Pract 2014;4(1):121-128. [doi: 10.5430/JNEP.V4N1P121]
21. Hollnagel E, Pàries J, Woods DD, Wreathall J. Resilience Engineering in Practice: A Guidebook. Farnham, UK: Ashgate Publishing; 2011.
22. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
23. Winwood PC, Colon R, McEwen K. A practical measure of workplace resilience: developing the resilience at work scale. J Occup Environ Med 2013 Oct;55(10):1205-1212. [doi: 10.1097/JOM.0b013e3182a2a60a] [Medline: 24064782]
24. Resilience test. Psychometric Test. 2013 Apr 29. URL: https://www.psychometrictest.org.uk/resilience-test/ [accessed 2022-05-04]

25. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
26. Connor KM, Davidson JR. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). Depress Anxiety 2003;18(2):76-82. [doi: 10.1002/da.10113] [Medline: 12964174]
27. Friborg O, Hjemdal O, Rosenvinge JH, Martinussen M. A new rating scale for adult resilience: what are the central protective resources behind healthy adjustment? Int J Methods Psychiatr Res 2003;12(2):65-76 [FREE Full text] [doi: 10.1002/mpr.143] [Medline: 12830300]
28. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. J Nurs Meas 1993;1(2):165-178. [Medline: 7850498]
29. Bartone PT, Ursano RJ, Wright KM, Ingraham LH. The impact of a military air disaster on the health of assistance workers. A prospective study. J Nerv Ment Dis 1989 Jun;177(6):317-328. [doi: 10.1097/00005053-198906000-00001] [Medline: 2723619]
30. Troy AS, Willroth EC, Shallcross AJ, Giuliani NR, Gross JJ, Mauss IB. Psychological resilience: an affect-regulation framework. Annu Rev Psychol 2023 Jan 18;74:547-576 [FREE Full text] [doi: 10.1146/annurev-psych-020122-041854] [Medline: 36103999]
31. Bowling J, Vercruysse C, Krinner LM, Greene T, Bello-Ogunu F, Webster C. A simultaneous concept analysis of resilience, coping, posttraumatic growth, and thriving. Nurs Forum 2022 Sep;57(5):905-919. [doi: 10.1111/nuf.12754] [Medline: 35643843]
32. Moya J, Goenechea M. An approach to the unified conceptualization, definition, and characterization of social resilience. Int J Environ Res Public Health 2022 May 09;19(9):5746 [FREE Full text] [doi: 10.3390/ijerph19095746] [Medline: 35565141]
33. Métais C, Burel N, Gillham JE, Tarquinio C, Martin-Krumm C. Integrative review of the recent literature on human resilience: from concepts, theories, and discussions towards a complex understanding. Eur J Psychol 2022 Feb 25;18(1):98-119 [FREE Full text] [doi: 10.5964/ejop.2251] [Medline: 35330859]
34. Daly LM. Resilience: an integrated review. Nurs Sci Q 2020 Oct;33(4):330-338. [doi: 10.1177/0894318420943141] [Medline: 32940160]
35. Tay PK, Lim KK. Psychological resilience as an emergent characteristic for well-being: a pragmatic view. Gerontology 2020;66(5):476-483. [doi: 10.1159/000509210] [Medline: 32784301]
36. Infurna FJ. What does resilience signify? An evaluation of concepts and directions for future research. Gerontology 2020;66(4):323-331. [doi: 10.1159/000507365] [Medline: 32408298]
37. Cooper AL, Brown JA, Rees CS, Leslie GD. Nurse resilience: a concept analysis. Int J Ment Health Nurs 2020 Aug;29(4):553-575. [doi: 10.1111/inm.12721] [Medline: 32227411]
38. Huey CW, Palaganas JC. What are the factors affecting resilience in health professionals? A synthesis of systematic reviews. Med Teach 2020 May;42(5):550-560. [doi: 10.1080/0142159X.2020.1714020] [Medline: 31984844]
39. Stainton A, Chisholm K, Kaiser N, Rosen M, Upthegrove R, Ruhrmann S, et al. Resilience as a multimodal dynamic process. Early Interv Psychiatry 2019 Aug;13(4):725-732. [doi: 10.1111/eip.12726] [Medline: 30126047]
40. Lyng HB, Macrae C, Guise V, Haraldseid-Driftland C, Fagerdal B, Schibevaag L, et al. Exploring the nature of adaptive capacity for resilience in healthcare across different healthcare contexts; a metasynthesis of narratives. Appl Ergon 2022 Oct;104:103810 [FREE Full text] [doi: 10.1016/j.apergo.2022.103810] [Medline: 35635941]
41. Raetze S, Duchek S, Maynard MT, Wohlgemuth M. Resilience in organization-related research: an integrative conceptual review across disciplines and levels of analysis. J Appl Psychol 2022 Jun;107(6):867-897. [doi: 10.1037/apl0000952] [Medline: 34766796]
42. Vercio C, Loo LK, Green M, Kim DI, Beck Dallaghan GL. Shifting focus from burnout and wellness toward individual and organizational resilience. Teach Learn Med 2021;33(5):568-576 [FREE Full text] [doi: 10.1080/10401334.2021.1879651] [Medline: 33588654]
43. Biddle L, Wahedi K, Bozorgmehr K. Health system resilience: a literature review of empirical research. Health Policy Plan 2020 Oct 01;35(8):1084-1109 [FREE Full text] [doi: 10.1093/heapol/czaa032] [Medline: 32529253]
44. Bhandari S, Alonge O. Measuring the resilience of health systems in low- and middle-income countries: a focus on community resilience. Health Res Policy Syst 2020 Jul 17;18(1):81 [FREE Full text] [doi: 10.1186/s12961-020-00594-w] [Medline: 32680527]
45. Strametz R, Raspe M, Ettl B, Huf W, Pitz A. Recommended actions: reinforcing clinicians' resilience and supporting second victims during the COVID-19 pandemic to maintain capacity in the healthcare system [Article in German]. Zentralbl Arbeitsmed Arbeitsschutz Ergon 2020;70(6):264-268 [FREE Full text] [doi: 10.1007/s40664-020-00405-7] [Medline: 32905028]
46. Wu AW, Connors C, Everly GSJ. COVID-19: peer support and crisis communication strategies to promote institutional resilience. Ann Intern Med 2020 Jun 16;172(12):822-823 [FREE Full text] [doi: 10.7326/M20-1236] [Medline: 32251512]
47. Thomas S, Sagan A, Larkin J, Cylus J, Figueras J, Karanikolos M. Strengthening Health Systems Resilience: Key Concepts and Strategies [Internet]. Copenhagen, Denmark: European Observatory on Health Systems and Policies; 2020.

48. Turenne CP, Gautier L, Degroote S, Guillard E, Chabrol F, Ridde V. Conceptual analysis of health systems resilience: a scoping review. Soc Sci Med 2019 Jul;232:168-180. [doi: 10.1016/j.socscimed.2019.04.020] [Medline: 31100697]
49. Qureshi A, Hargest C, Swain N, Aldabe D, Hale L. Psychosocial interventions for building resilience of informal carers of people living with stroke: a systematic review. Disabil Rehabil 2023 May;45(9):1419-1432. [doi: 10.1080/09638288.2022.2063419] [Medline: 35468030]
50. Stoliker BE, Vaughan AD, Collins J, Black M, Anderson GS. Building personal resilience following an online resilience training program for BScN students. West J Nurs Res 2022 Aug;44(8):755-764. [doi: 10.1177/01939459211017240] [Medline: 33998338]
51. Golden EA, Zweig M, Danieletto M, Landell K, Nadkarni G, Bottinger E, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: design process, distribution, and evaluation. JMIR Form Res 2021 May 05;5(5):e26590 [FREE Full text] [doi: 10.2196/26590] [Medline: 33872189]
52. Spiva L, Davis S, Case-Wirth J, Hedenstrom L, Hogue V, Box M, et al. The effectiveness of charge nurse training on leadership style and resiliency. J Nurs Adm 2020 Feb;50(2):95-103. [doi: 10.1097/NNA.0000000000000848] [Medline: 31977947]
53. Liu JJ, Ein N, Gervasio J, Battaion M, Reed M, Vickers K. Comprehensive meta-analysis of resilience interventions. Clin Psychol Rev 2020 Dec;82:101919. [doi: 10.1016/j.cpr.2020.101919] [Medline: 33045528]
54. Venegas CL, Nkangu MN, Duffy MC, Fergusson DA, Spilg EG. Interventions to improve resilience in physicians who have completed training: a systematic review. PLoS One 2019 Jan 17;14(1):e0210512 [FREE Full text] [doi: 10.1371/journal.pone.0210512] [Medline: 30653550]
55. Franco PL, Christie LM. Effectiveness of a one day self-compassion training for pediatric nurses' resilience. J Pediatr Nurs 2021;61:109-114. [doi: 10.1016/j.pedn.2021.03.020] [Medline: 33839602]
56. Donaldson SI, Donaldson SI. The positive functioning at work scale: psychometric assessment, validation, and measurement invariance. J Well Being Assess 2021 Jan 11;4:181-215. [doi: 10.1007/s41543-020-00033-1]
57. Butler J, Kern ML. The PERMA-Profiler: a brief multidimensional measure of flourishing. Int J Wellbeing 2016;6(3):1-48. [doi: 10.5502/ijw.v6i3.526]
58. Scott SD, Hirschinger LE, Cox KR, McCoig M, Hahn-Cover K, Epperly KM, et al. Caring for our own: deploying a systemwide second victim rapid response team. Jt Comm J Qual Patient Saf 2010 May;36(5):233-240. [doi: 10.1016/s1553-7250(10)36038-7] [Medline: 20480757]
59. Una guía para mandos intermedios: recomendaciones para reforzar la resiliencia de los profesionales, los equipos y las instituciones sanitarias. Equipo Resiliencia. URL: https://www.eresiliencia.es/es/ [accessed 2025-01-13]
60. Presentació eResiliencia. YouTube. URL: https://www.youtube.com/watch?v=8P_p3e6MyVU [accessed 2025-01-13]
61. Herrin S. Examining relationships among leadership behaviors, demographic and program factors, and resilience in physical therapy professional education program directors. J Phys Ther Educ 2023 Dec 01;37(4):342-350 [FREE Full text] [doi: 10.1097/JTE.0000000000000310] [Medline: 38478790]
62. Howe A, Smajdor A, Stöckl A. Towards an understanding of resilience and its relevance to medical training. Med Educ 2012 Apr;46(4):349-356. [doi: 10.1111/j.1365-2923.2011.04188.x] [Medline: 22429170]
63. Martin-Delgado J, Viteri E, Mula A, Serpa P, Pacheco G, Prada D, et al. Availability of personal protective equipment and diagnostic and treatment facilities for healthcare workers involved in COVID-19 care: a cross-sectional study in Brazil, Colombia, and Ecuador. PLoS One 2020 Nov 11;15(11):e0242185 [FREE Full text] [doi: 10.1371/journal.pone.0242185] [Medline: 33175877]
64. Martin-Delgado J, Poblete R, Serpa P, Mula A, Carrillo I, Fernández C, et al. Contributing factors for acute stress in healthcare workers caring for COVID-19 patients in Argentina, Chile, Colombia, and Ecuador. Sci Rep 2022 May 19;12(1):8496 [FREE Full text] [doi: 10.1038/s41598-022-12626-2] [Medline: 35589975]
65. Turan N, Canbulat Ş. The effectiveness of the training program on accepting and expressing emotions on the psychological resilience and depression levels of nurses: a two-year follow-up study. Arch Psychiatr Nurs 2023 Jun;44:1-7. [doi: 10.1016/j.apnu.2023.03.002] [Medline: 37197852]
66. Burton NW, Pakenham KI, Brown WJ. Feasibility and effectiveness of psychosocial resilience training: a pilot study of the READY program. Psychol Health Med 2010 May;15(3):266-277. [doi: 10.1080/13548501003758710] [Medline: 20480432]
67. Rogers D. Which educational interventions improve healthcare professionals' resilience? Med Teach 2016 Dec;38(12):1236-1241. [doi: 10.1080/0142159X.2016.1210111] [Medline: 27573430]
68. Henshall C, Ostinelli E, Harvey J, Davey Z, Aghanenu B, Cipriani A, et al. Examining the effectiveness of web-based interventions to enhance resilience in health care professionals: systematic review. JMIR Med Educ 2022 Sep 06;8(3):e34230 [FREE Full text] [doi: 10.2196/34230] [Medline: 36066962]
69. Mahaffey BL, Mackin DM, Rosen J, Schwartz RM, Taioli E, Gonzalez A. The disaster worker resiliency training program: a randomized clinical trial. Int Arch Occup Environ Health 2021 Jan;94(1):9-21 [FREE Full text] [doi: 10.1007/s00420-020-01552-3] [Medline: 32448931]

70. López-Pineda A, Carrillo I, Mula A, Guerra-Paiva S, Strametz R, Tella S, et al. Strategies for the psychological support of the healthcare workforce during the COVID-19 pandemic: the ERNST study. Int J Environ Res Public Health 2022 May 02;19(9):5529 [FREE Full text] [doi: 10.3390/ijerph19095529] [Medline: 35564924]
71. Mira JJ, Vicente MA, Lopez-Pineda A, Carrillo I, Guilabert M, Fernández C, et al. Preventing and addressing the stress reactions of health care workers caring for patients with COVID-19: development of a digital platform (Be + against COVID). JMIR Mhealth Uhealth 2020 Oct 05;8(10):e21692 [FREE Full text] [doi: 10.2196/21692] [Medline: 32936769]
72. Pacutova V, Geckova AM, de Winter AF, Reijneveld SA. Opportunities to strengthen resilience of health care workers regarding patient safety. BMC Health Serv Res 2023 Oct 19;23(1):1127 [FREE Full text] [doi: 10.1186/s12913-023-10054-0] [Medline: 37858175]
73. Smith AF, Plunkett E. People, systems and safety: resilience and excellence in healthcare practice. Anaesthesia 2019 Apr;74(4):508-517 [FREE Full text] [doi: 10.1111/anae.14519] [Medline: 30585298]
74. Aase K, Guise V, Billett S, Sollid SJ, Njå O, Røise O, et al. Resilience in Healthcare (RiH): a longitudinal research programme protocol. BMJ Open 2020 Oct 26;10(10):e038779 [FREE Full text] [doi: 10.1136/bmjopen-2020-038779] [Medline: 33109657]
75. Brown L, Cohen B, Costello R, Brazhnik O, Galis ZS. Next steps: operationalizing resilience research. Stress Health 2023 Sep;39(S1):62-66. [doi: 10.1002/smi.3256] [Medline: 37184110]
76. Ejaz H, Sultan B, Pienaar AJ, Froelicher ES. Effectiveness of a resilience-focused educational program for promoting resilience in nursing students: a systematic review and meta-analysis. Nurse Educ Pract 2024 Jul;78:104014. [doi: 10.1016/j.nepr.2024.104014] [Medline: 38879909]
77. Palacio GC, Krikorian A, Gómez-Romero MJ, Limonero JT. Resilience in caregivers: a systematic review. Am J Hosp Palliat Care 2020 Aug;37(8):648-658. [doi: 10.1177/1049909119893977] [Medline: 31830813]

Edited by A Kushniruk; submitted 07.10.24; peer-reviewed by D Bolivar, M Munikumar; comments to author 23.12.24; revised version received 30.12.24; accepted 31.12.24; published 05.02.25.

Please cite as:
Gil-Hernández E, Carrillo I, Martin-Delgado J, García-Torres D, Mira JJ
Development of a Web-Based Intervention for Middle Managers to Enhance Resilience at the Individual, Team, and Organizational Levels in Health Care Systems: Multiphase Study
JMIR Hum Factors 2025;12:e67263
URL: https://humanfactors.jmir.org/2025/1/e67263
doi:10.2196/67263
PMID:

Think-Aloud Testing of a Companion App for Colonoscopy Examinations: Usability Study

Christine Jacob[1], PhD; Roman Müller[1], MSc; Sonja Schüler[1], PhD; Alix Rey[2,3], PhD; Guillaume Rey[2], PhD; Berj Armenian[4*], MD; Alain Vonlaufen[4*], MD, PhD; Michael Drepper[4*], MD; Marius Zimmerli[5*], MD

[1]FHNW University of Applied Sciences and Arts Northwestern Switzerland, Riggenbachstrasse 16, Olten, Switzerland

[2]Gimini Biosciences SàRL, Geneva, Switzerland

[3]International Institute for Management Development, Lausanne, Switzerland

[4]GGHA Geneva Gastroenterology & Hepatology Associates SA, Geneva, Switzerland

[5]Clarunis University Digestive Health Care Center, Basel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Christine Jacob, PhD
FHNW University of Applied Sciences and Arts Northwestern Switzerland, Riggenbachstrasse 16, Olten, Switzerland

Abstract

Background: Colonoscopies are vital for initial screening, follow-ups, surveillance of neoplasia, and assessing symptoms such as rectal bleeding. Successful colonoscopies require thorough colon preparation, but up to 25% fail due to poor preparation. This can lead to longer procedures, repeat colonoscopies, inconvenience, poorer health outcomes, and higher costs. eHealth tools can enhance bowel preparation and potentially reduce the need for repeat procedures.

Objective: This usability study aimed to identify strengths and weaknesses in a prototype companion app for colonoscopy examinations. The objective was to obtain in-depth insights into the app's usability, ease of use, and content comprehension, with the aim of refining the tool to effectively fulfill its intended purpose, guided by feedback from potential users.

Methods: From February to August 2024, we conducted a qualitative study using the think-aloud procedure. Each session involved 6 tasks and a semistructured interview to delve deeper into participants' task experiences. All think-aloud sessions and interviews were recorded. Quantitative usability questions were analyzed using Microsoft Excel, while qualitative data underwent coding and analysis based on thematic analysis principles.

Results: In total, 17 individuals, all smartphone users, participated in this study. Participants were recruited from 1 hospital, 1 private clinic, and 1 patient organization in Switzerland. The study found that participants rated the app's usability metrics positively, with an overall mean rating of ease of use at 4.29 (SD 0.59), usefulness at 4.53 (SD 0.72), and comprehensibility at 4.29 (SD 0.92). For the individual features, the mean ratings for ease of use were between 4 and 4.65, usefulness ranged from 4.35 to 4.82, and comprehensibility received ratings between 4.29 and 4.53, all measured on a 5-point scale, where 1 represented low agreement and 5 indicated high agreement. Additionally, 100% of participants indicated they will or may use the app if they require a colonoscopy examination. Participants highlighted the need for reminders and alerts in the week leading up to the colonoscopy, along with tailored content, simplified language, and visual aids.

Conclusions: The app prototype demonstrated favorable results with the majority of participants, and the testing process enabled the prompt identification and resolution of usability issues. The next phase will prioritize and assess potential improvements based on urgency and feasibility to guide a focused development plan. Usability testing highlighted features such as push notifications and personalized content as top priorities for participants, making them key areas for immediate attention. Moving forward, the app has the potential to function effectively as a companion app for colonoscopy examinations. To achieve this, further studies with a larger sample in real-world settings will be crucial.

(JMIR Hum Factors 2025;12:e67043) doi:10.2196/67043

KEYWORDS

eHealth; mobile health; mHealth; digital health; technology assessment; technology adoption; technology implementation; usability study; colonoscopy; app; application; examinations; smartphone; usability

Introduction

Background

Colonoscopies are widely recognized as the most reliable method for detecting colorectal issues; their effectiveness and safety hinge on the thoroughness of bowel preparation beforehand [1,2]. Ensuring adequate bowel preparation is crucial for achieving clear visualization of the colon's inner lining during the procedure. Poor bowel preparation is linked to risks such as missed significant lesions, procedural challenges, longer operation times, higher rates of interval colorectal cancers, and increased health care expenses [3]. However, a colonoscopy is an invasive procedure that demands extensive preparation. This includes taking a laxative, restricting food and liquid intake, and stopping certain medications in the week before the appointment.

Research indicates that up to 11% of individuals miss their colonoscopy appointments [4,5], and among those who do attend, many have insufficient bowel preparation, hindering clear colonic visualization [6-12]. A recent study investigated the efficacy of various bowel preparation regimens (4L, 2L, and ≤1L) for colorectal cancer screening, focusing on key quality indicators such as bowel cleanliness, cecal intubation rate, adenoma detection rate, and polyp detection rate, all aligned with the performance standards set by the European Society of Gastrointestinal Endoscopy (ESGE) [13]. While all regimens met the ESGE's minimum quality thresholds, the adequacy of bowel preparation varied significantly between volumes [13]. Ultralow-volume preparations achieved an adequacy rate of 79%, notably lower than the 86.4% seen with high-volume preparations [13]. In particular, bowel preparation with sodium picosulfate and magnesium citrate (SPMC) and 1L polyethylene glycol with ascorbic acid (1L-PEGA) was adequate in only 75.2% and 82.9% of cases, respectively, highlighting the need for careful consideration when selecting a preparation method based on patient needs and procedural goals [13].

Misunderstanding dietary guidelines and cleansing instructions, along with noncompliance, significantly contributes to inadequate bowel preparation [14]. Ineffective bowel preparation can lead to several adverse outcomes, including reduced adenoma detection rates, extended procedure times, lower cecal intubation rates, increased electrocautery risks, and more frequent examination intervals [15,16]. To enhance patient adherence to colonoscopy procedures, various educational strategies have been used. Tools such as booklets, cartoons, and SMS text messaging have proven effective in increasing follow-up rates compared to standard care [17]. Furthermore, smartphone-based strategies have been developed to assist patients in preparing for colonoscopy [18-24]. Research indicates that these smartphone interventions generally lead to better outcomes, such as higher bowel cleansing quality scores, compared to usual care control groups [18-20,25,26]. However, there is limited evidence that these tools were designed with input from their intended users [27], which may reduce their effectiveness. Engaging potential users in the app development process is likely to enhance usability by ensuring the app's content and features match user needs and preferences [28-30].

Objectives

Considering that user research can significantly enhance a tool's adoption and adherence rates post launch [28], this usability study was conducted to pinpoint strengths and weaknesses in the prototype of a companion app for colonoscopy examinations and to provide detailed insights into its quality regarding usefulness, ease of use, and content comprehension. The goal is to refine the product based on evidence gathered from potential users, ensuring it meets its intended purpose effectively.

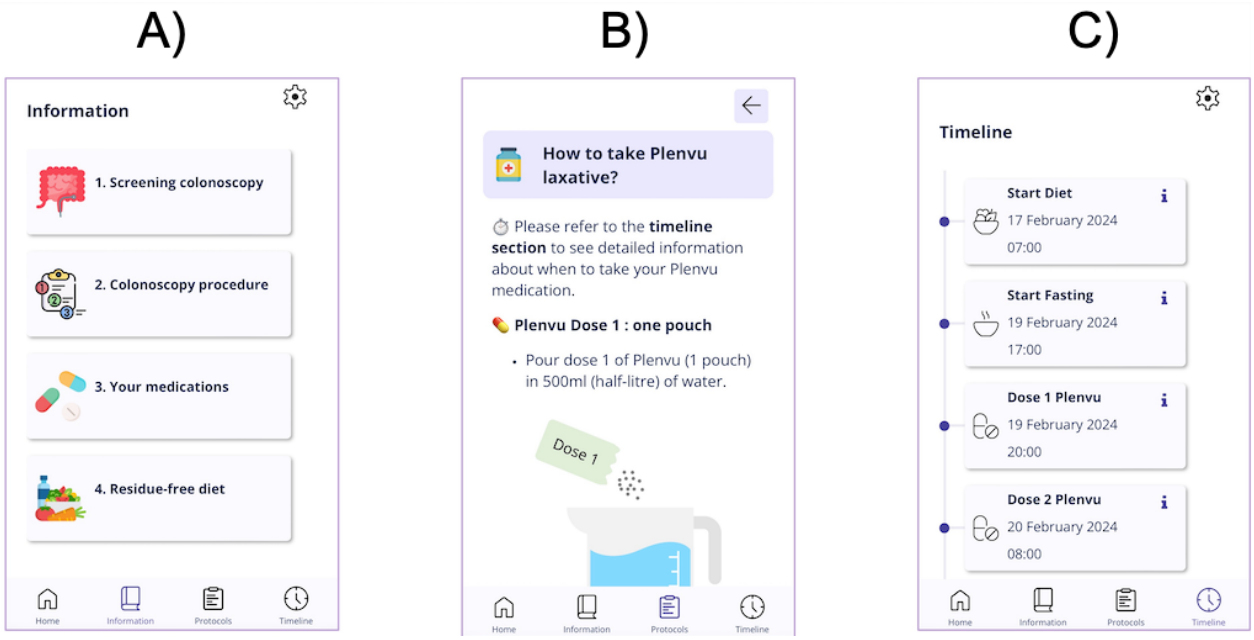
Methods

App Prototype

The health care technology company Gimini Biosciences SàRL is developing digital health solutions to empower patients undergoing complex medical interventions. Their mission is to ensure every patient has access to the necessary information for a successful medical examination.

The companion app for colonoscopy examinations, their first use case, features three main sections: (1) educational content about colonoscopy examinations; (2) digital protocols detailing diet, fasting, and laxative instructions; and (3) a personalized timeline based on the user's examination date and time to guide them on diet, fasting, and laxative schedules. Figure 1 illustrates the design of the prototype, showcasing these sections.

Figure 1. App prototype overview, including (A) patient information, (B) digital protocols, and (C) timeline and schedule.



Study Design

The study involved conducting a qualitative interview with each participant using the standardized think-aloud method [31], which is described in detail below. Participants were guided using a semistructured test script to articulate their cognitive process in real time while completing a list of predefined tasks. Similar to many qualitative studies, this research used purposive sampling to gather in-depth insights [32]. Participants were selected based on their capacity to offer detailed and firsthand insights into the research topic, ensuring they could effectively articulate their real-life experiences [32,33].

Participant Recruitment

The inclusion criteria covered individuals aged 40 - 65 years who use a smartphone, have undergone a colonoscopy examination, have access to Wi-Fi and email, are comfortable using teleconferencing tools (eg, Microsoft Teams), and are capable of screen-sharing during the testing session. Three institutions played a key role in recruiting participants: Clarunis (the University Digestive Health Care Center of St. Clara Hospital and University Hospital Basel), GGHA (Geneva Gastroenterology & Hepatology Associates SA), and EUPATI Switzerland (European Patients' Academy on Therapeutic Innovation).

Participants were directly approached and recruited by the collaborating institutions. Once a participant agreed to participate and signed the consent form, the respective institution forwarded the signed consent form along with the participant's contact details to the core research team at the University of Applied Sciences and Arts Northwest Switzerland (FHNW). From that point, the research team took over the coordination and management of the participant's involvement in the study.

Participants had the option to conduct the testing session in English, French, or German. The English version of the participant information sheet and consent form can be found in Multimedia Appendix 1. The information was also provided in both German and French, giving participants the freedom to select the language they felt most comfortable using, and the recruitment process spanned from February to July 2024. The researchers aimed to recruit a sufficient number of participants to achieve saturation, indicating that enough data had been collected when new information no longer provided additional insights [32,33].

Think-Aloud Procedure

The think-aloud method is extensively used in app development as a popular tool for assessing usability [34]. In a think-aloud test, participants are asked to use the system and verbalize their thoughts continuously as they navigate through the user interface. This technique allows researchers to gain deeper insights into user misconceptions, which often lead to practical recommendations for redesign: when users misunderstand design elements, those elements may require modification. It also sheds light on why users make incorrect assumptions about certain aspects of the design and why they find other parts intuitive. The method is used to make cognitive processes, which would otherwise remain implicit and unspoken, more visible. By having participants verbalize their thoughts while performing a task, researchers can gain insights into thought processes, decision-making strategies, and individual patterns of interpretation.

Participants evaluated a web-based prototype of the app. They were instructed to explore the prototype while vocalizing their actions and observations, and to offer feedback on features, navigation, and perceived usefulness as they interacted with it. Researchers reminded participants to maintain a continuous stream of thoughts when needed and observed their behavior throughout the test tasks. All sessions were recorded, capturing participants' screen interactions with the prototype. Detailed notes were taken during feedback sessions and stored using the research platform Tivian.

During each test session, participants were asked to complete 6 tasks, during which we observed their behavior and

interactions with the tool. Additionally, they provided subjective assessments for each task, evaluating its usefulness, ease of use, and comprehensibility of the content. Textbox 1 displays the list of tasks participants were required to complete during the test sessions. The English version of the complete test script can be found in Multimedia Appendix 2.

Textbox 1. The list of tasks participants were required to complete during the test sessions.

Task 1: Customization and language choice
Task 2: First screen tour
Task 3: Accessing and navigating the background information section
Task 4: Accessing and navigating the fasting and food instructions
Task 5: Accessing and navigating the laxative instructions
Task 6: Accessing and navigating the meal and exercise recommendations

The think-aloud test was designed to encompass every possible task a user might perform within the app, covering all available functions and interactions. By including all tasks, we ensured a thorough capture of user feedback across the app's entire functionality, providing a comprehensive assessment of the user experience. Below is a high-level summary of what each task involved:

- Task 1: Customization and language selection—In this task, participants were prompted to select their preferred language and adjust settings such as font size or content complexity, allowing them to personalize the app to their needs.
- Task 2: First screen tour—This task aimed to assess the intuitiveness of the app's initial screen, encouraging participants to reflect on the clarity of the navigation and their understanding of the content available in each section.
- Task 3: Exploring background information—Here, participants were guided to access and navigate the section on background information about the colonoscopy examination, supporting their understanding of the procedure as they prepared for their appointment.
- Task 4: Reviewing fasting and dietary instructions—This task directed participants to find and review guidelines on what they can and cannot eat before the examination, including details on when to begin the special diet and the required fasting period.
- Task 5: Reviewing laxative instructions—Participants were asked to locate and understand instructions on taking the laxative, including information on dosage timing for effective preparation.
- Task 6: Reviewing meal and exercise recommendations—This task involved participants exploring the app's suggestions for meals and exercises that could help them prepare optimally for the examination.

Analysis

After compiling all notes from the testing sessions, CJ, RM, and SS collectively reviewed and synthesized the findings. These were then presented to the app development team for discussion, focusing on identifying features and functions requiring modification. The data encompassed audio recordings of the think-aloud sessions, observations noted by CJ, RM, and SS, and responses to usability questions.

Quantitative usability data, such as user ratings of ease of use and usefulness of the different features, were aggregated and analyzed using Microsoft Excel for Mac 2021 (version 16.86) to compute totals, percentages, means, and standard deviations. Qualitative participant comments were translated into English as needed for coding purposes. NVivo version 1.7.2 (QSR International), a qualitative data analysis software, was used for coding and categorizing the qualitative data. The data underwent thematic analysis to capture the depth and the unique interpretative contributions of individual researchers [35]. To ensure coding reliability, the first 3 authors, who conducted testing sessions in 3 different languages, engaged in collaborative discussions. The initial codebook was organized around core eHealth app development components: user interface (including navigation and visual design), user experience design (such as personalization), functionality (including notifications and reminders), and patient engagement and support (such as educational content). Initial coding was carried out by the first author (CJ), followed by a review from RM. Any coding discrepancies were addressed through discussions with SS until a consensus was reached.

Ethical Considerations

The ethics committee of Northwest and Central Switzerland determined that ethics approval was not needed for this study, according to the Federal Act on Research Involving Human Beings, article 2, paragraph 1 (reference number Req-2023 - 01506). All participants were briefed about the research background and signed a consent form agreeing to participate. Participants did not receive payment but were offered the opportunity for early and free access to the app upon its launch.

Results

Sample Characteristics

In total, 17 participants from 4 institutions tested the app prototype (3 of them were pilot tests, 1 in each of the 3 test languages). Participants were mostly male (13/17, 76%) and aged 40 - 50 (8/17, 47%) years. The gender discrepancy among participants is largely due to recruitment challenges and the voluntary nature of the study. Since participation was optional, it may have led to a self-selection bias, where individuals more comfortable with technology or who have a specific interest in health applications were more likely to take part, resulting in a less balanced gender distribution among participants. Tests were conducted in German (6/17, 35%), French (6/17, 35%), and

English (5/17, 29%). Table 1 presents the demographics and characteristics of the sample.

Table . Sample characteristics (N=17).

Characteristic		Values, n (%)
Gender		
	Male	13 (76)
	Female	4 (24)
Age (years)		
	<40	1 (6)[a]
	40 - 50	8 (47)
	51 - 60	5 (29)
	>60	3 (18)
Language		
	German	6 (35)
	French	6 (35)
	English	5 (29)
Referring institution		
	Clarunis[b]	9 (53)
	GGHA[c]	4 (24)
	FHNW[d] (pilot testers)	3 (18)
	EUPATI Switzerland[e]	1 (6)

[a]One of the pilot tests, hence age inclusion criteria were not applied.

[b]Clarunis: The University Digestive Health Care Center of St. Clara Hospital and University Hospital Basel.

[c]GGHA: Geneva Gastroenterology & Hepatology Associates SA.

[d]FHNW: University of Applied Sciences and Arts, Northwestern Switzerland.

[e]EUPATI Switzerland: European Patients' Academy on Therapeutic Innovation, Switzerland.

Usability Metrics

Usability metrics are specific measurements used to evaluate a digital product's usability. These metrics typically assess how quickly users complete tasks, how often they make mistakes, and their overall satisfaction with the tool. By analyzing various usability metrics, we can gain a comprehensive understanding of the user's experience and the tool's overall usability.

During the testing sessions, we incorporated satisfaction metrics, which are subjective measures based on users' self-assessments. These metrics evaluated the ease of use, usefulness, and content comprehensibility of different sections of the app, as well as their overall impression of the app. Responses were rated on a Likert scale from 1 (low agreement) to 5 (high agreement). Table 2 provides an overview of these subjective measures for the various tasks and the app as a whole. Some measures were not applicable for certain tasks. For example, in task 1 (customization and language selection), there was no text content to assess for comprehensibility, so this measure is marked as N/A (not applicable) for that task. In task 2 (first screen tour), participants were only asked to reflect on whether the initial screen was intuitive, rather than completing an action. Therefore, measures such as ease of use, usefulness, and comprehensibility were not rated for this task.

Table . Subjective usability measures (N=17).

Task	Ease of use[a], mean (SD)	Usefulness[a], mean (SD)	Comprehensibility[a], mean (SD)
Task 1: Customization and language choice	4.41 (0.71)	4.35 (1.11)	N/A[b]
Task 2: First screen tour	N/A	N/A	N/A
Task 3: Accessing and navigating the background information section	4.41 (0.94)	4.76 (0.44)	4.53 (0.87)
Task 4: Accessing and navigating the fasting and food instructions	4 (1)	4.41 (0.87)	4.29 (0.99)
Task 5: Accessing and navigating the laxative instructions	4.53 (0.62)	4.82 (0.39)	4.35 (0.93)
Task 6: Accessing and navigating the meal and exercise recommendations	4.65 (0.7)	4.59 (0.62)	4.53 (0.8)
Overall satisfaction with the app as a whole	4.29 (0.59)	4.53 (0.72)	4.29 (0.92)

[a]User satisfaction of usability attributes was rated on a scale of 1 (low agreement) to 5 (high agreement).

[b]N/A: not applicable.

To provide a comprehensive assessment, test moderators also recorded observations on three additional usability metrics: (1) completion metrics (these measure whether users can successfully complete or partially complete tasks, indicating the tool's effectiveness); (2) duration metrics (these track the average time users take to perform a task, reflecting the design's complexity and the efficiency of user navigation); and (3) error metrics (these refer to actions users take that do not lead to the expected outcome, highlighting areas of confusion in the user interface or challenges with functionality). Table 3 presents an overview of these observed measures. Certain measures in this table were not applicable for task 2 (first screen tour), as participants were only asked to assess the intuitiveness of the initial screen rather than perform any specific action. As a result, metrics such as completion rate and error rate were not relevant for this task, since there was no actionable step for participants to complete or errors to quantify.

Table . Observed usability measures (N=17).

Task	Completion, n (%)	Error rate, n (%)	Duration (seconds), mean (SD)
Task 1: Customization and language choice	16 (94)	1 (6)	44.47 (32.39)
Task 2: First screen tour	N/A[a]	N/A	57.41 (45.94)
Task 3: Accessing and navigating the background information section	15 (88)	7 (41)	66.53 (46.32)
Task 4: Accessing and navigating the fasting and food instructions	16 (94)	7 (41)	75.41 (46.07)
Task 5: Accessing and navigating the laxative instructions	15 (88)	3 (18)	54.88 (47.89)
Task 6: Accessing and navigating the meal and exercise recommendations	17 (100)	3 (18)	56.76 (50.49)

[a]N/A: not applicable.

When asked if they would use the app for a colonoscopy examination once it becomes available, out of 17 participants, 14 (82%) said yes, 3 (18%) said maybe, and 0 (0%) said no.

Qualitative Feedback

Participants provided comments and qualitative feedback on the 6 tasks they performed during the test sessions. While their feedback concerning the comprehensibility of the content, the user-friendliness, and the general user benefits of the application was mostly positive, they also expressed some confusion about certain content or features. Additionally, they offered suggestions for improvements in areas such as design and visualization as well as user guidance to address the gaps or issues they identified in the prototype. Overall, participants stressed the importance of receiving reminders and alerts in the week leading up to their colonoscopy. They also preferred tailored content, simplified language, and visual aids to enhance their understanding. Table 4 summarizes their qualitative feedback, organized into 4 key themes: app content, design, guidance, and features. These themes collectively illustrate the

participants' preferences and priorities, providing valuable insights for the development and improvement of the application.

Table . Key themes that emerged from the thematic analysis of the qualitative feedback.

Themes and subthemes		Qualitative feedback
Content		
	Content accuracy	Participants expressed a need for greater accuracy in the app's content. Specifically, they suggested: • Allowing the selection of examination times down to the minute, rather than in 15-minute intervals. • Clearly stating that tea and hot drinks should not contain milk, as specified in the paper-based guidance. • Ensuring the app accurately reflects the strict fasting phase, as some participants experienced longer fasting periods than indicated.
	Content clarity	Some content has been found to be unclear or confusing. For instance: • The term "residue-free diet" may be difficult for laypeople to understand; "bowel cleansing" is suggested as a more straightforward alternative. • The term "protocols" is ambiguous, and "preparation" is recommended as a clearer option. • Instead of using the word "option" for meal examples, provide explicit examples to avoid misleading users into thinking these are the only choices available. • For dosage instructions, emphasize that the liquid should be consumed in sips rather than all at once. • The timeline entry "pickup medication" is unclear and potentially misleading; a more descriptive label is needed. • More information on physical exercise should be included, such as its effects and importance in the preparation process.
	Content completeness	It has been noted that certain content is currently lacking in the app, namely: • Expanded details on the colonoscopy procedure should be provided, with careful attention to wording. Starting the information section with cancer detection details can be perceived as alarming. It is essential to introduce the procedure in a reassuring manner, highlighting its benefits and emphasizing preventive care. • Information on the effects of laxatives. • Vegetarian menu options. • Post-colonoscopy care and potential adverse events. • Any assistance with laxative consumption would be greatly appreciated, as it can be quite unpleasant. For example, mixing the laxative with clear syrup to improve taste. Another tip is to sip water alternately with the laxative instead of consuming them sequentially, which can make the experience less difficult. Additional advice on making the process more manageable could include using a straw to drink the laxative or drinking tea beforehand to mitigate the salty taste.

Themes and subthemes		Qualitative feedback
	FAQs[a]	Participants have suggested including a section with FAQs to provide clear and concise answers to common concerns. Suggested questions include: • Why is a colonoscopy performed? Include an explanation of its purpose and benefits, and note any differences in the procedure or considerations for men and women. • Can I drive after taking the medication? • When can I resume a normal diet? • What are the advantages and disadvantages of various examination methods? Offer a comparison of different diagnostic options, such as stool examinations, to help users understand their relative benefits and limitations.
	Translation accuracy	A few minor translation inaccuracies have been identified: • “Dose 1” was incorrectly translated into German; the correct term is “Dosis.” • Under “fasting time,” “Untersuchung” is a more appropriate translation than “Prüfung.”
Design		

Themes and subthemes		Qualitative feedback
	Navigation clarity	Participants have proposed several improvements to enhance the clarity of the app's navigation. Suggested enhancements include: • Adjust the color scheme of the language selection navigation to make it more prominent. • The button for confirming language selection should be available in multiple languages. • Change the name of the navigation section labeled "Information" to better reflect that it pertains to the procedure rather than app-related information. • The back arrow should be made more prominent by increasing its size and repositioning it centrally for better visibility and ease of use. • There is uncertainty about the type of information the sections labeled "Protocols" and "Timeline" contain.
	Navigation structure	There were several suggestions to improve the app's navigation structure: • Increase the size of the customization icon and consider positioning it at the same level as the other main navigation icons to attract more attention. • Keep content concise and allow users to click for more detailed information when needed. • Display the timeline immediately after scheduling an appointment, as it is highly relevant. Ideally, place the timeline in the same section where laxative information is provided. • Clearly mark the section for taking the laxative in the protocol and integrate the "how" with the "when" to provide a comprehensive guide. • Distinguish and prioritize personalized content related to individual preparation for the colonoscopy from general informational content about the procedure, such as examination details and menu suggestions.
Guidance		

Themes and subthemes		Qualitative feedback
	Step-by-step guidance	Participants have reported some confusion and expressed a need for clearer guidance on navigating the app and understanding required actions. Specific recommendations include: • Add a note indicating that users can change the language and the date of their intervention, as this option is not immediately apparent. • Clearly explain why users need to enter their appointment details. • Provide a clear explanation of the next steps immediately after entering appointment information, as some users were unsure where to click. • Indicate that some content is customized based on the user's appointment time to clarify how the app personalizes information. • Include preliminary information about the general procedure at the start of the app, such as an overview page, to give users a better understanding of what to expect.
Features		
	Personalized notifications	Nearly all participants emphasized the critical need for personalized reminders and alerts, tailored to each patient based on the specific date and time of their colonoscopy appointment.
	Sharing and printing info	The importance of having the ability to print or share information directly from the app was emphasized.
	Tailored content	Participants noted that the app's capability to customize content for individual users makes it more favorable compared to other methods of information delivery, such as paper. Specific suggestions include: • Set the app to automatically select the language based on the user's browser settings. • Allow direct integration with the user's calendar for streamlined timeline management. • Ensure that the link directs users to the correct hospital right from the start. • Add the medical center's phone number or provide a direct link for users to easily contact the center.
	Visual aids	Participants recommended integrating visual aids to improve the clarity and comprehension of the information. Suggestions include: • Add illustrations to assist in understanding the content more easily. • Use visual cues to highlight the fasting time more prominently. • Present information about meal preparation and dishes in a more visual format, including photos of menu items. • Integrate photos and potentially a video to visually explain the procedure.

[a]FAQs: frequently asked questions.

Discussion

Principal Findings and Implications for App Improvement

Our findings indicate that participants were generally willing to use the companion app for colonoscopy examinations and found it mostly useful and easy to navigate. Usability is crucial for implementation, as technologies that are user-friendly are more likely to be consistently used over time. This is why leading technology acceptance frameworks such as the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT) emphasize ease of use and usefulness as key predictors for adopting new technologies [36,37]. The inclusion of subjective measures in assessments has been debated in the literature due to potential variability introduced by users' subjective views. Despite this challenge, many scholars advocate for the inclusion of subjective criteria, such as ease of use and visual appeal, because they are fundamental drivers of adoption [38-41]. Therefore, incorporating subjective criteria, such as perceived ease of use, into the testing process could enhance tool adherence and improve health outcomes [42]. However, it is important to recognize that moderating factors such as a person's age, education, and digital skills can influence their assessment of a technology's perceived ease of use and overall usefulness [38].

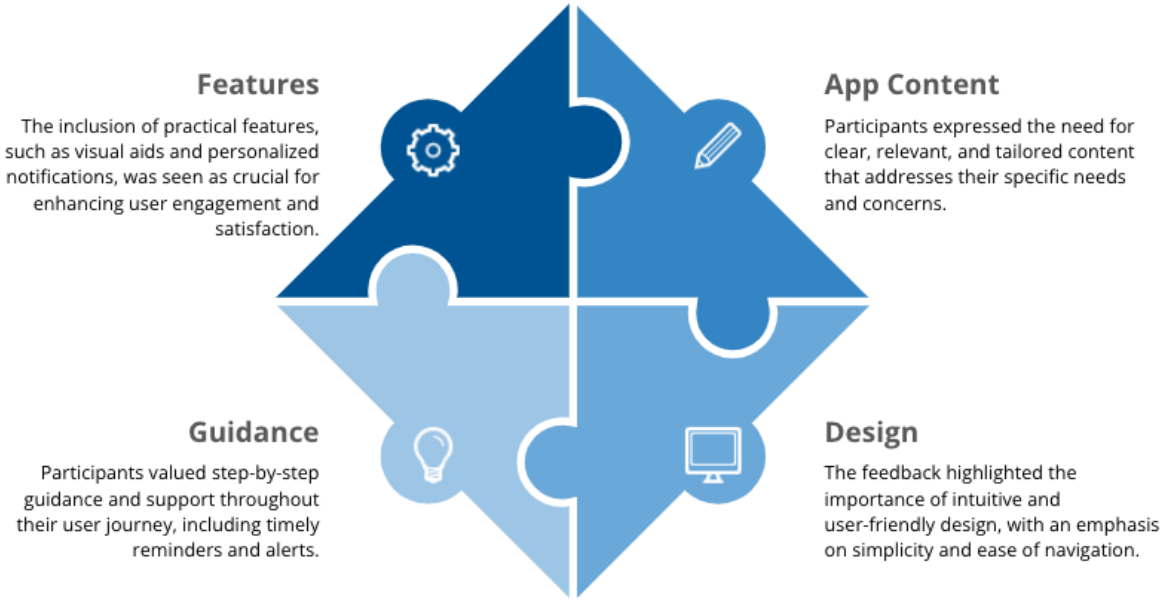
Although participants rated the app highly for both ease of use and usefulness, they also offered several suggestions for improvement. For instance, they recognized that preparing for a colonoscopy involves multiple complex steps during the week leading up to the appointment. They expressed a desire for the app to send timely reminders to guide users throughout the process, aligning with findings from previous similar research [27]. Reminders are likely crucial for the effectiveness of the app. In fact, a randomized controlled trial found that without reminders the app group had the same quality of bowel cleanliness as the control group [21]. Previous studies have demonstrated that personalized reminders and alerts enhance the adoption of electronic health technologies, effectively modifying various health behaviors [43].

The app content was generally well received for its clarity and ease of understanding. However, there were a few requests for improved clarity or accuracy in certain areas, as detailed in the Results section. Additionally, users identified some content gaps in the tested prototype. Specifically, they requested more comprehensive guidance on how to take the laxative and minimize the unpleasantness of the experience. Furthermore, there was a need for information addressing post-colonoscopy issues. This latter finding aligns with similar research; for instance, Sewitch et al [27] also highlighted the perceived importance of post-colonoscopy information among potential users. Their study similarly underscored the need for detailed guidance on what to expect and how to manage any subsequent issues after the procedure. This is especially pertinent considering that statistics reveal approximately 25% of patients experience a minor adverse event within 48 hours of a colonoscopy, and 0.5% encounter a serious adverse event within 30 days [44].

Participants indicated a need for more detailed, step-by-step guidance on using the app. They sought clear instructions on the use of entering specific information, such as their procedure date and time, and how to effectively use the app to prepare for their examination. This feedback aligns with the extensive literature emphasizing the importance of user training and guidance to enhance adoption rates [38]. Introducing an app tour with guided notes detailing each section, coupled with clear instructions on features such as adjusting font size, could greatly enhance the overall user experience. This initial guidance would make it easier for users to understand how to navigate the app and maximize its use right from the start. Research suggests that users' perception of ease of use can be significantly improved with high-quality training materials that guide them on how to effectively optimize these technologies [38,45,46].

Participants expressed a strong preference for more visual and interactive content, as well as a navigation design that prioritizes personalized information, such as the preparation timeline. They expressed a desire for information to be presented clearly at every stage of the colonoscopy preparation process. This feedback is consistent with previous research, which underscores the critical role of design elements and personalization in achieving successful patient adoption of digital health tools [38]. Several studies have specifically highlighted the importance of personalization. For example, a lack of customization options to meet individual needs can result in lower adoption rates or even abandonment of the tool [38,47]. Figure 2 provides a summary of our findings regarding user preferences for app design, content, features, and guidance. These insights will help boost the app's usability and functionality, ultimately improving user adoption.

Figure 2. Key areas for app improvement.



The next step will involve a thorough prioritization and impact assessment, where improvements according to urgency and feasibility will be ranked to guide a targeted development plan. For example, the usability testing results indicate that features such as push notifications and personalized content are top priorities for the participants. These will be addressed at the start of the next development cycle to ensure they are optimized before moving into real-world testing.

Limitations and Future Research

While the practical insights from participants regarding the app prototype were valuable and guided us in making important modifications, it is important to acknowledge several limitations. Using the think-aloud methodology, there is a potential for reporting bias where participants may shape their responses to align with perceived researcher expectations; additionally, participants might selectively verbalize thoughts that align with what they believe interviewers want to hear, known as social desirability bias [48]. To address these potential biases, moderators were trained to actively encourage participants to provide verbal feedback and to collect observational data. These observations were used alongside participants' subjective feedback during data analysis.

While the sample size of this study may appear small, previous research indicates that 80% - 90% of usability issues in websites and apps can be identified with 5 to 9 participants [49,50]. During the individual think-aloud sessions, recurring themes emerged, suggesting data saturation was achieved [32]. However, it is possible that additional participants could have uncovered different usability issues and provided diverse perspectives. Looking ahead, the app has promising potential as a companion for colonoscopy examinations. However, conducting larger-scale studies in real-world environments will be essential to validate and optimize its value and effectiveness.

Conclusions

Overall, participants expressed satisfaction with the app's usability. The think-aloud sessions provided real-time insights into the app's appeal, relevance, and use. Minor adjustments to the prototype's functionality were identified as necessary to enhance usability. Feedback and suggestions from participants have been integrated into the final app design. The initial findings from this usability study indicate that the app holds promising potential as a companion for colonoscopy examinations. This work establishes the foundation for further research to evaluate usability and feasibility among a larger, real-world population.

Acknowledgments

This study was sponsored by Innosuisse (Swiss Innovation Agency, grant 70325.1 INNO-ICT). The role of Innosuisse is to promote science-based innovation in the interests of industry and society in Switzerland. We extend our gratitude to the 3 institutions that were instrumental in recruiting the study participants: Clarunis (the University Digestive Health Care Center of St. Clara Hospital and University Hospital Basel), GGHA (Geneva Gastroenterology & Hepatology Associates SA), and EUPATI Switzerland (European Patients' Academy on Therapeutic Innovation).

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. Restrictions may apply to the availability of some of these data due to, for example, privacy and ethical reasons.

Conflicts of Interest

AR and GR are cofounders of Gimini Biosciences SàRL, the technology provider developing the application being studied. 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
Participant information sheet and consent form.
[PDF File, 139 KB - humanfactors_v12i1e67043_app1.pdf]

Multimedia Appendix 2
Think-aloud usability test script.
[PDF File, 244 KB - humanfactors_v12i1e67043_app2.pdf]

References

1. Beck DE. Bowel preparation for colonoscopy. Clin Colon Rectal Surg 2010 Feb;23(1):10-13. [doi: 10.1055/s-0030-1247851] [Medline: 21286285]
2. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014 Apr 3;370(14):1298-1306. [doi: 10.1056/NEJMoa1309086] [Medline: 24693890]
3. Lebwohl B, Kastrinos F, Glick M, Rosenbaum AJ, Wang T, Neugut AI. The impact of suboptimal bowel preparation on adenoma miss rates and the factors associated with early repeat colonoscopy. Gastrointest Endosc 2011 Jun;73(6):1207-1214. [doi: 10.1016/j.gie.2011.01.051] [Medline: 21481857]
4. Dalton ARH. Incomplete diagnostic follow-up after a positive colorectal cancer screening test: a systematic review. J Public Health 2018;40(1):e46-e58. [doi: 10.1093/pubmed/fdw147]
5. Chopra D, Hookey LC. Comorbid illness, bowel preparation, and logistical constraints are key reasons for outpatient colonoscopy nonattendance. Can J Gastroenterol Hepatol 2016;2016:2179354. [doi: 10.1155/2016/2179354] [Medline: 27478818]
6. Sulz MC, Kröger A, Prakash M, Manser CN, Heinrich H, Misselwitz B. Meta-analysis of the effect of bowel preparation on adenoma detection: early adenomas affected stronger than advanced adenomas. PLoS One 2016;11(6):e0154149. [doi: 10.1371/journal.pone.0154149] [Medline: 27257916]
7. Govani SM, Elliott EE, Menees SB, et al. Predictors of suboptimal bowel preparation in asymptomatic patients undergoing average-risk screening colonoscopy. World J Gastrointest Endosc 2016 Sep 16;8(17):616-622. [doi: 10.4253/wjge.v8.i17.616] [Medline: 27668072]
8. Yee R, Manoharan S, Hall C, Hayashi A. Optimizing bowel preparation for colonoscopy: what are the predictors of an inadequate preparation? Am J Surg 2015 May;209(5):787-792. [doi: 10.1016/j.amjsurg.2014.12.018] [Medline: 25796095]
9. Dik VK, Moons LMG, Hüyük M, et al. Predicting inadequate bowel preparation for colonoscopy in participants receiving split-dose bowel preparation: development and validation of a prediction score. Gastrointest Endosc 2015 Mar;81(3):665-672. [doi: 10.1016/j.gie.2014.09.066] [Medline: 25600879]
10. Kherad O, Restellini S, Martel M, Barkun AN. Polyethylene glycol versus sodium picosulfalte bowel preparation in the setting of a colorectal cancer screening program. Can J Gastroenterol Hepatol 2015 Oct;29(7):384-390. [doi: 10.1155/2015/350587] [Medline: 26301330]
11. Menees SB, Kim HM, Wren P, et al. Patient compliance and suboptimal bowel preparation with split-dose bowel regimen in average-risk screening colonoscopy. Gastrointest Endosc 2014 May;79(5):811-820. [doi: 10.1016/j.gie.2014.01.024] [Medline: 24631492]
12. Kurlander JE, Sondhi AR, Waljee AK, et al. How efficacious are patient education interventions to improve bowel preparation for colonoscopy? A systematic review. PLoS One 2016;11(10):e0164442. [doi: 10.1371/journal.pone.0164442] [Medline: 27741260]
13. Theunissen F, Lantinga MA, Ter Borg PCJ, et al. Efficacy of different bowel preparation regimen volumes for colorectal cancer screening and compliance with European Society of Gastrointestinal Endoscopy performance measures. United European Gastroenterol J 2023 Jun;11(5):448-457. [doi: 10.1002/ueg2.12386] [Medline: 37190897]
14. Froehlich F, Wietlisbach V, Gonvers JJ, Burnand B, Vader JP. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. Gastrointest Endosc 2005 Mar;61(3):378-384. [doi: 10.1016/s0016-5107(04)02776-2] [Medline: 15758907]
15. Alvi H, Rasheed T, Shaikh MA, Ali FS, Zuberi BF, Samejo AA. Impact of bowel preparation on caecal intubation time during colonoscopy. Pak J Med Sci 2019;35(6):1516-1519. [doi: 10.12669/pjms.35.6.1031] [Medline: 31777485]
16. Kluge MA, Williams JL, Wu CK, et al. Inadequate Boston Bowel Preparation Scale scores predict the risk of missed neoplasia on the next colonoscopy. Gastrointest Endosc 2018 Mar;87(3):744-751. [doi: 10.1016/j.gie.2017.06.012] [Medline: 28648575]

17. Liu Z, Zhang MM, Li YY, Li LX, Li YQ. Enhanced education for bowel preparation before colonoscopy: a state-of-the-art review. J Dig Dis 2017 Feb;18(2):84-91. [doi: 10.1111/1751-2980.12446] [Medline: 28067463]
18. Cho J, Lee S, Shin JA, Kim JH, Lee HS. The impact of patient education with a smartphone application on the quality of bowel preparation for screening colonoscopy. Clin Endosc 2017 Sep;50(5):479-485. [doi: 10.5946/ce.2017.025] [Medline: 28669148]
19. Walter B, Schmid R, von Delius S. A smartphone app for improvement of colonoscopy preparation (ColoprAPP): development and feasibility study. JMIR Mhealth Uhealth 2017 Sep 20;5(9):e138. [doi: 10.2196/mhealth.7703] [Medline: 28931498]
20. Walter BM, Klare P, Neu B, Schmid RM, von Delius S. Development and testing of an automated 4-day text messaging guidance as an aid for improving colonoscopy preparation. JMIR Mhealth Uhealth 2016;4(2):e75. [doi: 10.2196/mhealth.5289]
21. Sharara AI, Chalhoub JM, Beydoun M, et al. A customized mobile application in colonoscopy preparation: a randomized controlled trial. Clin Transl Gastroenterol 2017 Jan 5;8(1):e211. [doi: 10.1038/ctg.2016.65] [Medline: 28055031]
22. Jung JW, Park J, Jeon GJ, et al. The effectiveness of personalized bowel preparation using a smartphone camera application: a randomized pilot study. Gastroenterol Res Pract 2017;2017:4898914. [doi: 10.1155/2017/4898914] [Medline: 28928768]
23. Kang X, Zhao L, Leung F, et al. Delivery of instructions via mobile social media app increases quality of bowel preparation. Clin Gastroenterol Hepatol 2016 Mar;14(3):429-435. [doi: 10.1016/j.cgh.2015.09.038] [Medline: 26492848]
24. Back SY, Kim HG, Ahn EM, et al. Impact of patient audiovisual re-education via a smartphone on the quality of bowel preparation before colonoscopy: a single-blinded randomized study. Gastrointest Endosc 2018 Mar;87(3):789-799. [doi: 10.1016/j.gie.2017.09.007]
25. van der Zander QEW, Reumkens A, van de Valk B, Winkens B, Masclee AAM, de Ridder RJJ. Effects of a personalized smartphone app on bowel preparation quality: randomized controlled trial. JMIR Mhealth Uhealth 2021 Aug 19;9(8):e26703. [doi: 10.2196/26703] [Medline: 34420924]
26. Aksan F, Tanriverdi LH, Figueredo CJ, Barrera LC, Hasham A, Jariwala SP. The impact of smartphone applications on bowel preparation, compliance with appointments, cost-effectiveness, and patients' quality of life for the colonoscopy process: a scoping review. Saudi J Gastroenterol 2023;29(2):71-87. [doi: 10.4103/sjg.sjg_207_22] [Medline: 36073572]
27. Sewitch MJ, Fallone CA, Ghali P, Lee GE. What patients want in a smartphone app that supports colonoscopy preparation: qualitative study to inform a user-centered smartphone app. JMIR Mhealth Uhealth 2019 Jul 2;7(7):e12242. [doi: 10.2196/12242] [Medline: 31125310]
28. Jacob C, Bourke S, Heuss S. From testers to cocreators—the value of and approaches to successful patient engagement in the development of eHealth solutions: qualitative expert interview study. JMIR Hum Factors 2022 Oct 6;9(4):e41481. [doi: 10.2196/41481] [Medline: 36102548]
29. Aguayo GA, Goetzinger C, Scibilia R, et al. Methods to generate innovative research ideas and improve patient and public involvement in modern epidemiological research: review, patient viewpoint, and guidelines for implementation of a digital cohort study. J Med Internet Res 2021 Dec 23;23(12):e25743. [doi: 10.2196/25743] [Medline: 34941554]
30. Harte R, Glynn L, Rodríguez-Molinero A, et al. A human-centered design methodology to enhance the usability, human factors, and user experience of connected health systems: a three-phase methodology. JMIR Hum Factors 2017;4(1):e8. [doi: 10.2196/humanfactors.5443]
31. Ericsson KA, Simon HA. How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. Mind Cult Act 1998 Jul;5(3):178-186. [doi: 10.1207/s15327884mca0503_3]
32. Braun V, Clarke V. Successful Qualitative Research: A Practical Guide for Beginners: SAGE Publications; 2013.
33. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches: SAGE Publications; 2017.
34. Nielsen J. Thinking aloud: the #1 usability tool. Nielsen Norman Group. 2012 Jan 15. URL: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ [accessed 2025-01-17]
35. Braun V, Clarke V. What can "thematic analysis" offer health and wellbeing researchers? Int J Qual Stud Health Well-being 2014;9(1):26152. [doi: 10.3402/qhw.v9.26152] [Medline: 25326092]
36. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
37. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
38. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR Mhealth Uhealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
39. Jacob C, Lindeque J, Klein A, Ivory C, Heuss S, Peter MK. Assessing the quality and impact of eHealth tools: systematic literature review and narrative synthesis. JMIR Hum Factors 2023;10:e45143. [doi: 10.2196/45143]
40. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. BMJ Open 2021 Mar 19;11(3):e047001. [doi: 10.1136/bmjopen-2020-047001] [Medline: 33741674]
41. Jacob C, Lindeque J, Müller R, et al. A sociotechnical framework to assess patient-facing eHealth tools: results of a modified Delphi process. NPJ Digit Med 2023 Dec 15;6(1):232. [doi: 10.1038/s41746-023-00982-w] [Medline: 38102323]

42. Uncovska M, Freitag B, Meister S, Fehring L. Rating analysis and BERTopic modeling of consumer versus regulated mHealth app reviews in Germany. NPJ Digit Med 2023 Jun 21;6(1):115. [doi: 10.1038/s41746-023-00862-3] [Medline: 37344556]
43. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
44. Sewitch MJ, Azalgara VM, Sing MF. Screening indication associated with lower likelihood of minor adverse events in patients undergoing outpatient colonoscopy. Gastroenterol Nurs 2018;41(2):159-164. [doi: 10.1097/SGA.0000000000000308] [Medline: 29596130]
45. Louissaint J, Lok AS, Fortune BE, Tapper EB. Acceptance and use of a smartphone application in cirrhosis. Liver Int 2020 Jul;40(7):1556-1563. [doi: 10.1111/liv.14494] [Medline: 32343467]
46. Blonigen D, Harris-Olenak B, Kuhn E, Humphreys K, Timko C, Dulin P. From "step away" to "stand down": tailoring a smartphone app for self-management of hazardous drinking for veterans. JMIR Mhealth Uhealth 2020 Feb 13;8(2):e16062. [doi: 10.2196/16062] [Medline: 32053118]
47. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020 Sep 29;22(9):e20283. [doi: 10.2196/20283] [Medline: 32990635]
48. Güss CD. What is going through your mind? Thinking aloud as a method in cross-cultural psychology. Front Psychol 2018;9:1292. [doi: 10.3389/fpsyg.2018.01292] [Medline: 30150948]
49. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/BF03195514]
50. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]

Abbreviations

1L-PEGA: 1L polyethylene glycol with ascorbic acid
Clarunis: the University Digestive Health Care Center of St. Clara Hospital and University Hospital Basel
ESGE: European Society of Gastrointestinal Endoscopy
EUPATI Switzerland: European Patients' Academy on Therapeutic Innovation
FHNW: University of Applied Sciences and Arts Northwest Switzerland
GGHA: Geneva Gastroenterology & Hepatology Associates SA
SPMC: sodium picosulfate and magnesium citrate
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 30.09.24; peer-reviewed by LAK Breuch, MM Kessler; revised version received 08.12.24; accepted 08.12.24; published 12.02.25.

Please cite as:
Jacob C, Müller R, Schüler S, Rey A, Rey G, Armenian B, Vonlaufen A, Drepper M, Zimmerli M
Think-Aloud Testing of a Companion App for Colonoscopy Examinations: Usability Study
JMIR Hum Factors 2025;12:e67043
URL: https://humanfactors.jmir.org/2025/1/e67043
doi:10.2196/67043

Comparison of a Novel Machine Learning–Based Clinical Query Platform With Traditional Guideline Searches for Hospital Emergencies: Prospective Pilot Study of User Experience and Time Efficiency

Hamza Ejaz[1], MD; Hon Lung Keith Tsui[2*], MD; Mehul Patel[2], MD; Luis Rafael Ulloa Paredes[2], MSc; Ellen Knights[1], BSc; Shah Bakht Aftab[3*], MD; Christian Peter Subbe[4*], Dr med

[1]Ysbyty Gwynedd, Clinical School, Clinical Research Office, Bangor, United Kingdom

[2]Medwise.ai Ltd, Leeds, United Kingdom

[3]Russells Hall Hospital, Dudley, United Kingdom

[4]North Wales Medical School, Bangor University, Brigantia Building, Bangor, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Christian Peter Subbe, Dr med
North Wales Medical School, Bangor University, Brigantia Building, Bangor, United Kingdom

Abstract

Background: Emergency and acute medicine doctors require easily accessible evidence-based information to safely manage a wide range of clinical presentations. The inability to find evidence-based local guidelines on the trust's intranet leads to information retrieval from the World Wide Web. Artificial intelligence (AI) has the potential to make evidence-based information retrieval faster and easier.

Objective: The aim of the study is to conduct a time-motion analysis, comparing cohorts of junior doctors using (1) an AI-supported search engine versus (2) the traditional hospital intranet. The study also aims to examine the impact of the AI-supported search engine on the duration of searches and workflow when seeking answers to clinical queries at the point of care.

Methods: This pre- and postobservational study was conducted in 2 phases. In the first phase, clinical information searches by 10 doctors caring for acutely unwell patients in acute medicine were observed during 10 working days. Based on these findings and input from a focus group of 14 clinicians, an AI-supported, context-sensitive search engine was implemented. In the second phase, clinical practice was observed for 10 doctors for an additional 10 working days using the new search engine.

Results: The hospital intranet group (n=10) had a median of 23 months of clinical experience, while the AI-supported search engine group (n=10) had a median of 54 months. Participants using the AI-supported engine conducted fewer searches. User satisfaction and query resolution rates were similar between the 2 phases. Searches with the AI-supported engine took 43 seconds longer on average. Clinicians rated the new app with a favorable Net Promoter Score of 20.

Conclusions: We report a successful feasibility pilot of an AI-driven search engine for clinical guidelines. Further development of the engine including the incorporation of large language models might improve accuracy and speed. More research is required to establish clinical impact in different user groups. Focusing on new staff at beginning of their post might be the most suitable study design.

(JMIR Hum Factors 2025;12:e52358) doi:10.2196/52358

KEYWORDS

artificial intelligence; machine learning; information search; emergency care; developing; testing; information retrieval; hospital care; training; clinical practice; clinical experience; user satisfaction; clinical impact; user group; users; study design; mobile phone

Introduction

In making decisions about patient care, clinicians frequently are faced with queries and are often unable to retrieve answers to them in a timely fashion. A systematic review [1] estimates that the per-patient frequency of queries raised by clinicians ranges from 0.4 to 0.8 per patient and that about two-thirds of these queries are left unanswered. This picture has been fairly stable over time despite the broad availability of web-based evidence resources that can answer these questions. Unanswered

questions may lead to suboptimal patient care decisions and are missed opportunities for timely learning and practice improvement.

Medwise.ai is a solution codeveloped with Betsi Cadwaladr University Health Board), which helps clinicians find answers quickly from local guidelines and provide seamless and just-in-time access to high-quality evidence in the context of patient care decision-making in the clinical environment. A previous study [2] has demonstrated that smartphone apps can increase the speed to access guidelines when compared to using desktop computers. However, access to guidelines does not equate to finding answers to clinical questions quickly and effectively at the point of care. The Medwise.ai platform can collate local clinical guidelines and break them down into chunks of content that can be retrieved using natural language processing and information retrieval technologies as bite-sized answers for clinician's questions. We conducted a time-motion analysis, comparing cohorts of junior doctors using an artificial intelligence (AI)–supported search engine and the traditional hospital intranet, to examine the impact of the AI-supported search engine on participant time and workflow when seeking answers to clinical queries at the point of care. The secondary end points of our study were to assess whether effective question answering could lead to a better quality of work life and improve confidence in decision-making for the end user.

This study aimed (1) to gain user feedback for a novel context-specific proof-of-concept search engine and (2) to measure the time required to retrieve clinically relevant information with a novel engine compared to other search engines in the setting of hospital emergency care.

Methods

Study Design

We conducted a prospective direct pre- and postobservational pilot study in teams caring for medical emergency admissions examining the impact of access to Medwise.ai, a clinical query answering platform, versus information retrieval from guidelines saved on the hospital intranet using time-motion study methodology [3].

Study Setting

The study took place at the Ysbyty Gwynedd, Bangor. The Ysbyty Gwynedd is a district general hospital in North Wales with 550 beds covering all major specialties, including a dialysis unit and a 13-bedded critical care unit. Clinicians from 2 units in the hospital were recruited; the acute medical unit has 23 beds and is complemented by a same day emergency care unit for low-risk admissions.

Both units receive direct admissions from primary care and work closely with the emergency department (ED) to manage acute medical presentations. It is important to note that the traditional boundaries between emergency care and acute medicine have become less distinct in recent years due to increasing hospital overcrowding in the United Kingdom. While historically, patients were first seen in the ED by emergency physicians and then referred to the Acute Medical Unit for care by acute and general physicians, many acute and general medical teams now also work within the ED to manage the flow of acute medical patients.

During the day shift (8 AM to 8:30 PM), the on-call team consists of 4 doctors in training (a newly qualified foundation year 1 doctor, 2 core medical trainees with 1‐4 years of experience, and 1 medical registrar with 4 or more years of clinical experience and membership of the Royal College of Physicians) and an on-call consultant with a full specialist qualification. The sole task of the on-call team is the care of emergency admissions and emergencies of inpatients. Patients seen by doctors in training are subsequently reviewed by a consultant as part of the posttake ward round.

Intervention

Medwise.ai is a proof-of-concept search engine combining well-established information retrieval techniques with textual question answering and trained models that determine the best answer within a document. The search platform is available to the participating clinicians over a web app accessible via mobile web browser on both Android and iOS devices. Research staff trained the onboard clinicians and helped to install the Medwise.ai search on the participating clinicians' mobile devices ahead of observations.

Content development for Medwise.ai was informed by informal interviews (Multimedia Appendix 1) with staff and a single focus group, consisting of 14 participants, to identify relevant local guidelines and standard operating procedure documents to be included in the Medwise.ai platform, thus ensuring user buy-in and optimal functionality of the Medwise.ai platform. The local documents shared with Medwise.ai were in PDF or Microsoft Word format. It is important to note that searches using Medwise.ai were limited by the content available in the local repository; if the content did not exist within this repository, the AI would not be able to provide answers to queries. It should be noted that the underlying AI model's technical performance was not the focus of this study.

Research Procedure

A dedicated and trained member of the research team shadowed trainee doctors for 20 complete working days: observation days included 10 days without and 10 days with access to Medwise.ai. The 2 groups of participants were not matched for the hospital intranet versus AI-supported search engine using groups.

Doctors reviewed new admissions and conducted unscheduled reviews of previously admitted patients: the doctors were aware that they were observed, and written consent was taken prior to commencing the direct observation. The observed activity of doctors was entered into a work diary. For practical reasons, all observations were undertaken during office hours.

Sampling and Recruitment

The sample size was chosen pragmatically based on experience with previous time-motion studies [4]. Doctors working as part of the acute medical take throughout Ysbyty Gwynedd were recruited. To be included, participants had to be willing and able to give informed consent for participation in the study, be in possession of a smartphone to access web-based content, be permanent or locum staff, and be qualified as a doctor of any

grade. Participants had the right to withdraw from the study at any time. Participants received vouchers with a value of US $50 for participation in the study. Doctors who were only part of the clinical team for 4 hours or less were excluded from recruitment.

Candidates were approached by the principal investigator (CPS) or the dedicated research team member. Notification of the study was undertaken through existing WhatsApp groups and a Junior Doctors' Forum. Inclusion and exclusion criteria were identified during the initial screening.

Assessments

Participants' gender, specialty, grade, duration of work in their current post, and duration of clinical practice were recorded. Participants were shadowed for the duration of a shift but for a minimum of 6 hours per study day. Shadowing did not include times when doctors were consulting or examining patients.

Task duration for clinical queries was measured in minutes. Tasks were classified using a standardized list. Observed searches for information were classified by clinical content (diagnostic category, diagnostic pathway, treatment protocols, prognostication, scoring systems, normal values, and accessed data source). Confidence in decision-making was assessed at the end of each working day using a validated scale for self-assessment [5]. Confidence in decision-making and user satisfaction were measured using the Likert scale (How satisfied are you with the Medwise.ai product? On a Likert scale of 1-10) and the Net Promoter Score (NPS) [6] (How likely are you to recommend Medwise.ai to a friend or colleague?). NPS is a validated score that is used to measure user satisfaction. It is widely used in industry as a benchmark for comparing products within different industries. A positive NPS is considered as good, a score above 20 as favorable, and above 50 as excellent. Participants were also asked for suggestions for improvements of the platform.

Data Management and Statistics

Statistical analysis was primarily descriptive including key data items related to the frequency of clinical searches, the subject of such searches, the time taken for the searches, and graded feedback in relation to the app. Data were recorded on paper case report forms and subsequently transcribed into a Microsoft Excel database for further analysis. Data were anonymized through the removal of name, case report number, and age.

The comparison was made between the events during 10 observed shifts without the app and 10 observed shifts with the app. Diary data and data from questionnaires were analyzed using SPSS (IBM Corp). Kernel density plots were used to visualize the differences in observed task duration; Welch *t* test (2-tailed) was used to determine whether the apparent differences in distribution were likely to have occurred due to random chance or were due to a real difference in the mean task duration.

Given the sample size and the unknown baseline distribution of the variables in question, statistical comparison data from this study might inform subsequent power calculations for related research or a definitive trial.

Ethical Considerations

National Health Service (NHS) Research Ethics Review was not required as the study involved staff only as participants. The study was approved through Health and Care Research Wales and the Health Research Authority. No adverse events were anticipated. However, monitoring was in place in line with the sponsor safety reporting standard operating procedure. Data from focus group meetings were anonymized and deidentified. Written informed consent was given by all members of staff who were observed during the assessments. Participants were reimbursed with £50 (US $61.92) retail vouchers according to local regulations.

Results

Demographics and Baseline Characteristics

All assessments were undertaken between November 17, 2022, and February 8, 2023 (hospital intranet period) and July 6 and 27, 2023 (AI-supported search engine period).

The hospital intranet and AI-supported search engine groups both comprised 10 doctors each with 5 and 7 female doctors, respectively. The hospital intranet group had a median clinical experience of 23 months. Of these, 6.1 months were in their current specialty. The AI-supported search engine group had a median of 54 months of experience. Of these, 9.6 months were in their current specialty. No statistically significant differences between mean time in current specialty were observed between the groups (t=−1.01; 95% CI −11.4679 to 4.0179; P>.05).

Focus Group

The focus group was held on April 5, 2023, after the completion of the hospital intranet period. In total, 14 clinicians participated including foundation year 1 and 2 doctors, registrars, physician associates, and 3 consultant physicians. Participants were asked a list of prepared questions in relation to search habits, preferred information sources, and views about search engines (Multimedia Appendix 1).

Contributors discussed sources of information that they use for their clinical practice and recommended examples of web resources from other NHS organizations. Pros and cons of alternative models of information provision including share point sites were debated. Clinicians commented on the difficulties to find the right locally authorized resources in locations of the internet and intranet. On the other hand, they valued that Medwise.ai guidance is relevant for the organization in which the search is undertaken. Participants discussed their experience of using resources from other UK and international organizations if there is no identifiable local guidance for a topic. Doctors were concerned that Google searches might identify sources that contain misinformation. The challenge of identifying the right information in real time was described as crucial during the time pressures of acute and emergency care.

Descriptives

Overall, 67 searches were observed in the hospital intranet and 39 searches in the AI-supported search engine period. The mean number of clinical searches performed per shift was 6.7 (SD 3.43) in the hospital intranet group and 3.9 (SD 3.00) in the

AI-supported search engine group. Mobile phones were used for 40 of 67 (60%) searches in the hospital intranet group and 34 of 39 (85%) in the AI-supported search engine group. In the hospital intranet group, 11 searches were via Google. Other sources included apps, the British National Formulary, National Institute for Health and Care Excellence guidelines, and local guidelines. In the AI-supported search engine group, no search required Google.

Searches covered a large number of topics (Multimedia Appendices 2 and 3). The most commonly searched topics in both groups were medication-related (n=18 and n=10, respectively).

Participants in the AI-supported search engine group commented on the impact of Medwise.ai on workflow using free-text responses (Textbox 1).

Textbox 1. Free-text feedback provided by users collected at the end of observation.

"With more training [training of the AI system] this could be very helpful in speeding up carrying out jobs. Bit unfortunate study was carried out in final weeks of rotation, when doctors are the most comfortable with systems and need to search things very rarely" [Participant 1].

"Can see it has the potential to increase speed of doing jobs, thus means we see patients quicker" [Participant 2].

"May expediate time available for patient care if brings up more precise answers. When I know where to find the answers it is easier to go straight to the source" [Participant 3].

"As you get your answers under one second from hospital protocol, it saves time and helps you see more patients. also can provide more time for patient care rather than looking for protocol and guidelines in each app, it will be all in one. Highly recommended" [Participant 4].

"Probably not much difference. Perhaps better used in GP. I think it would be more useful in GP. In hospital it is okay but not majorly useful. I think over all it would be a good idea. But I think I search for stuff more in GP. I think overall ED staff would use it most in hospital" [Participant 5].

"Easy availability of resources saves up time and causes less disruption to the flow of clerking. Easy access to BCU/local guidelines is very helpful as some of these are difficult to find on the intranet. Great idea; would be really helpful" [Participant 6].

"Will reduce search time for searching guidelines resulting in more time for patient care. And all in one place makes it easy to find [information]. I am surprised that I have not needed to access my phone/ the app as yet today as this is unusual. Would usually look at betsinet multiple times during a shift for electrolyte levels. I anticipate that this app would be useful" [Participant 7].

"Makes it easier to search for relevant guidelines/policies/pathways etc. Easy and quick to use. User friendly. Appears to work well for searches during work time. Quickly found everything I searched for" [Participant 8].

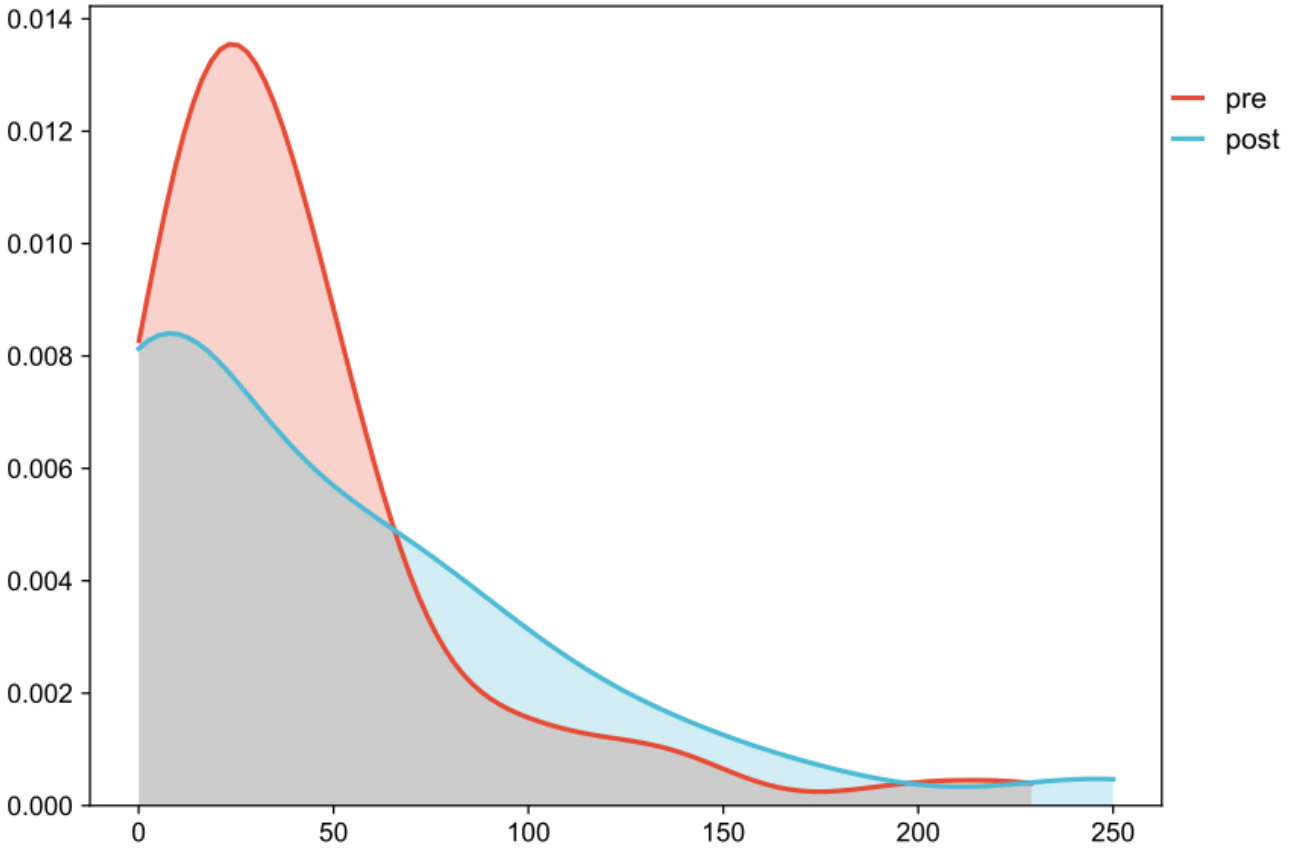
"Will help find relevant guidelines much more quickly, which will make work more efficient" [Participant 9].

"Directed me to mdcalc which I already use as an app. No impact on above criteria (workflow, number of patients seen and time for each patient). Would be more useful to have easy access to trust guidelines such as electrolyte imbalances and chest pain etc." [Participant 10].

Comparison of Groups

Comparison of the 2 groups revealed that searches in the hospital intranet group had shorter search times than in the AI-supported search engine group (41.4 vs 88.1 seconds, respectively; Welch t=4.06; 95% CI 20.30-59.09; P<0.05). The differences in the observed task duration are presented as Kernel density plots in Figure 1. There were no statistically significant differences between the groups with regard to user satisfaction (Welch t=0.75; 95% CI −1.2607 to 2.6607; P>.05) and likelihood of a search result solving the query (χ^2=2.2; 95% CI −3.24% to 27.45%; P>.05). Participants of the AI-supported search engine group were likely to recommend Medwise.ai to a friend or colleague as reflected in their NPS of 20 (40% promoters and 20% detractors).

Figure 1. Kernel plot of the duration of searches pre (hospital intranet) and post (artificial intelligence–supported search engine) phase.



Overall, the group using Medwise.ai spent longer searching for information, and their searches were not more likely to be successful compared to the usual information retrieval methods.

Discussion

What We Have Shown

To the best of our knowledge, this is the first study to evaluate the impact of an AI-based information retrieval system on clinical workflow in acute hospital care. The design and implementation of a clinical information search engine integrating intranet- and internet-based information resources using machine learning was feasible and well-received by participants.

Searches in this small sample were not faster, and there was no higher likelihood of a successful search result. Despite this, doctors believe Medwise.ai has the potential to improve efficiency and workflow in the future and hence highly recommended the digital tool to their colleagues.

Strength and Weaknesses

One of the strengths of our study is that it uses objective measures of efficiency including search times and search outcomes rather than relying on self-reported questionnaire data from users. This mitigates recall bias. Additionally, the free-text feedback from users provided useful insights in that it shows that despite Medwise.ai's limitations, the users were optimistic about its potential in the future and expressed views in favor of its implementation in clinical workflow, as evident in quotes in Textbox 1.

It is worth noting that no specific training in evidence searching was provided to the participants beyond what they would have received as part of their usual medical education. This lack of additional training could be seen as both a strength and a limitation of our study. On one hand, it reflects the real-world scenario where clinicians often rely on their existing search skills when using new tools. This approach allows for a more authentic assessment of the AI-supported search engine's usability and effectiveness in a clinical setting. On the other hand, it raises questions about whether targeted training in evidence searching or in using the new platform might have improved search efficiency and outcomes. The ideal search engine should require minimal introduction or training, but understanding its functionality could potentially enhance search yields. Future studies might consider comparing the performance of users with and without specific search training to better understand the impact of such preparation on the effectiveness of AI-supported search tools in clinical practice.

The initially unfamiliar user interface could have been a significant contributor to the longer search time. Additionally, the introduction of the AI-supported search engine occurred toward the end of the rotation when doctors were most confident about the processes and management of patients. Participants had spent an average of 8 months in their current specialty, throughout which they would have regularly used traditional information retrieval sources.

Our observations and user feedback highlighted several areas for potential improvement in future iterations of the AI-supported search engine. First, we recognized that the platform was initially developed for desktop use, but mobile

access emerged as the preferred method in the acute care setting. In response, the platform's developers are working on a mobile-first version, including dedicated iOS and Android apps, as well as access via a chatbot interface on WhatsApp. This adaptation aims to improve user experience and reduce search times. Additionally, we noted that the more experienced intervention group might have been less accustomed to new digital formats, potentially impacting their interaction with the novel interface. To address this, future iterations could incorporate more intuitive design elements and provide brief, targeted training to familiarize users with the system's capabilities.

In our pilot, Medwise.ai was neither more nor less efficient, successful, or satisfying to users than traditional information retrieval sources. Searches using Medwise.ai were limited by the content available: if content did not exist within the local repository, then the AI will not be able to provide answers to queries. Finally, the local documents that were shared with Medwise.ai were in PDF or Microsoft Word format. Further development to transform these documents into more machine-readable formats with seamless integration with the search could further improve the performance of the Medwise.ai search engine. Like all AI-based tools, Medwise.ai's performance is contingent upon the quality and quantity of data used to train the tool. As the bank of searches builds up over time, Medwise.ai would be able to offer easier and faster access to reliable local guidelines.

The effectiveness of the search engine is dependent on the underlying AI models' performance. Although not a focus of this study, it is an area for future attention, particularly with the growing use of large language models.

What Others Have Shown

The NHS Long Term Workforce Plan [7] emphasizes the need to leverage the power of AI to improve efficiency and workflow. This case study is an example of a technological innovation to improve work efficiency with the end goal of improving patient care. It contributes to the literature focusing on evaluating the potential of AI-based technologies to improve service provision in the NHS [8-10].

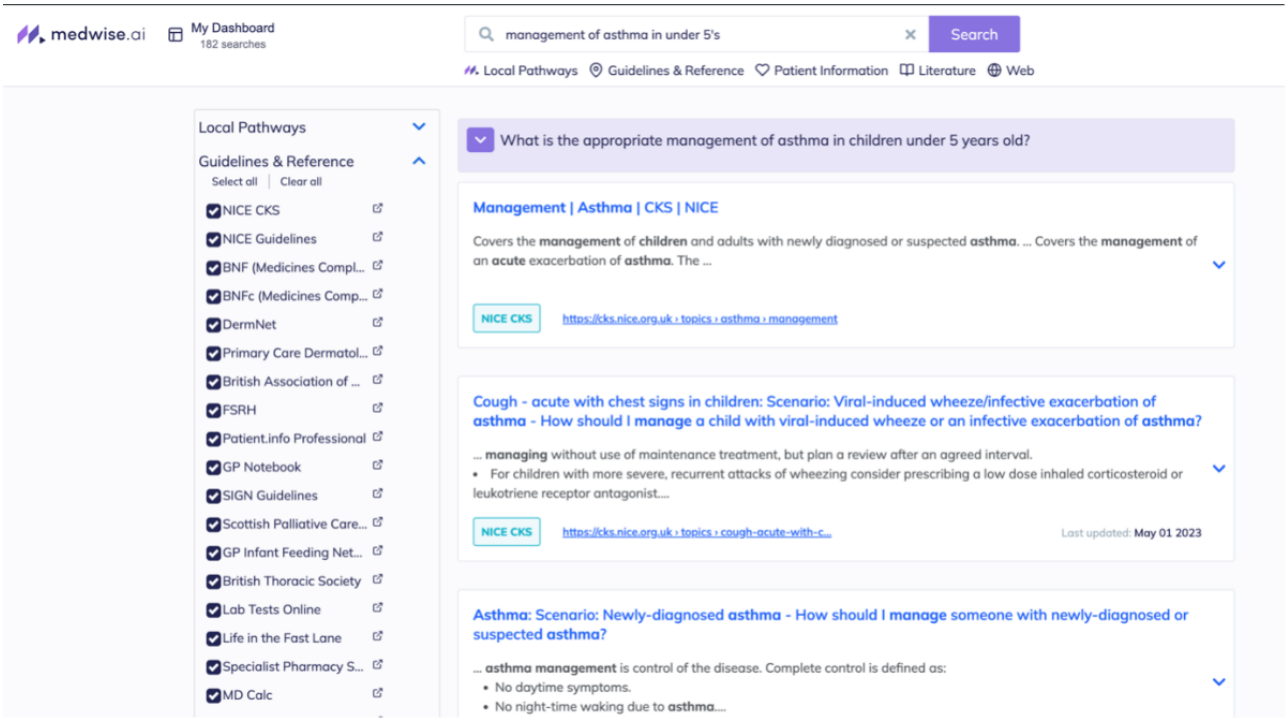
Clinical Implications

The search engine was implemented successfully during this trial. With the clear hierarchy of information sources, Medwise.ai might represent a source that has higher utility for clinicians at the bedside than public search engines such as Google or Safari. While it has been hypothesized that AI could aid to bring clinical guidelines to routine consultations [11], clinical apps remain scarce [12]. Moreover, evidence suggesting a significant reliance on colleagues and internet websites for information retrieval during shifts makes a strong case for the need for easy-to-access evidence-based local guidelines [13].

Research Recommendations

Given the small sample size and the unknown baseline distribution of the variables in question, statistical comparison is expected to be of limited value. A larger study with randomization of doctors at the beginning of rotational posts might provide a more robust assessment. The authors believe that less experienced clinicians (foundation year 1 trainees or advanced nurse practitioners in their first year of practice) would rely heavily on digital information sources and would be equally unfamiliar with other information retrieval methods and might hence be the most suitable group to evaluate effects on decision-making, safety, and possible clinical impact. More recent developments in AI such as large language models have shown the ability to encode clinical knowledge and answer medical questions [14]. Integrating these into the Medwise.ai engine might improve accuracy and speed. In the proof-of-concept platform, a full-text, query-augmented, retrieval engine configured and tuned for medical searches was used (Figure 2). No large language models were used. Further research is needed to evaluate how the accuracy and performance of the proof-of-concept search engine could be improved by incorporating large language models into the engine.

Figure 2. An image of the platform's search reply.



Conclusions

Implementation of Medwise.ai was feasible. In this small pilot study of an AI-supported search engine, we did not demonstrate increased workflow efficiency, search success, or satisfaction when compared with searches using the traditional hospital intranet. However, doctors in training believed the solution has the potential to do so in the future and hence recommend its implementation in clinical workflow. With added source content and integration of the Medwise.ai search with the next generation of AI large language models, it is likely that more benefits will be realized.

Acknowledgments

Marie Latham-Jones and Lynne Grundy were instrumental in establishing the connection between Betsi Cadwaladr University Health Board and Medwise.ai. Laura Longshaw was the driving force behind the progress of this grant. Without her encyclopedic knowledge of research frameworks, forensic attention to detail, unmeasurable patience, and immense commitment, the authors would not have succeeded. Without Gareth Jones' contribution to the recruitment of participants in the artificial intelligence–supported search engine group, the study would not have been completed on time. A big thank you! The study was funded through an Innovate UK grant administered through Medwise with Betsi Cadwaladr University Health Board as a coapplicant (Innovate UK grant application 10023176).

Conflicts of Interest

HLKT, LRUP, and MP are employees of Medwise AI Limited. The authors have no other conflict of interest.

Multimedia Appendix 1
Questions asked in the focus group.
[DOCX File, 16 KB - humanfactors_v12i1e52358_app1.docx]

Multimedia Appendix 2
Clustered column chart showing the distribution of search categories in the hospital intranet period (control) and artificial intelligence–supported search engine period (intervention). "General calc" comprised of calculating medication doses and converting units of measurements. "Interpreting invx" meant finding information about the significance of specific findings on investigations including plain films, electrocardiographs, and blood results. "Medication" meant medication-related information, that is, dose, route of administration, and interactions. "Scoring systems" meant users calculating prognostic scores including Well score. "Staff contact" meant looking up bleep numbers for other members of the health care workforce. "Term" meant looking up specific definitions. "Treatment protocol" meant treatment pathways for conditions and presenting complaints.
[JPEG File, 15 KB - humanfactors_v12i1e52358_app2.jpeg]

Multimedia Appendix 3
Word cloud of search terms.
[JPEG File, 132 KB - humanfactors_v12i1e52358_app3.jpeg]

References

1. Del Fiol G, Workman TE, Gorman PN. Clinical questions raised by clinicians at the point of care: a systematic review. JAMA Intern Med 2014 May;174(5):710-718. [doi: 10.1001/jamainternmed.2014.368] [Medline: 24663331]
2. Penders R, Mallet M. Survive On Call—a QI project to improve access to hospital clinical guidelines. BMJ Qual Improv Rep 2016;5(1):u210274.w4363. [doi: 10.1136/bmjquality.u210274.w4363] [Medline: 27239310]
3. Lopetegui M, Yen PY, Lai A, Jeffries J, Embi P, Payne P. Time motion studies in healthcare: what are we talking about? J Biomed Inform 2014 Jun;49:292-299. [doi: 10.1016/j.jbi.2014.02.017] [Medline: 24607863]
4. Sabin J, Khan W, Subbe CP, et al. "The time it takes…" How doctors spend their time admitting a patient during the acute medical take. Clin Med (Lond) 2016 Aug;16(4):320-324. [doi: 10.7861/clinmedicine.16-4-320] [Medline: 27481373]
5. Nagendran M, Chen Y. Real-time confidence of clinical decision making: a systematic review. Futur Healthc J 2017;17(2):82-90. [doi: 10.7861/futurehosp.6-2s-s82]
6. What is net promoter? NICE. URL: https://www.netpromoter.com/know/ [accessed 2022-06-03]
7. NHS Long Term Workforce Plan. NHS England. URL: https://www.england.nhs.uk/publication/nhs-long-term-workforce-plan/ [accessed 2025-01-03]
8. Akkus Z, Cai J, Boonrod A, et al. A survey of deep-learning applications in ultrasound: artificial intelligence-powered ultrasound for improving clinical workflow. J Am Coll Radiol 2019 Sep;16(9 Pt B):1318-1328. [doi: 10.1016/j.jacr.2019.06.004] [Medline: 31492410]
9. Bizzo BC, Almeida RR, Michalski MH, Alkasab TK. Artificial intelligence and clinical decision support for radiologists and referring providers. J Am Coll Radiol 2019 Sep;16(9 Pt B):1351-1356. [doi: 10.1016/j.jacr.2019.06.010] [Medline: 31492414]
10. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. N Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
11. Manning CL. Artificial intelligence could bring relevant guidelines into every consultation. BMJ 2019 Aug 7;366:l4788. [doi: 10.1136/bmj.l4788] [Medline: 31391184]
12. Hamed E, Sharif A, Eid A, Alfehaidi A, Alberry M. Advancing artificial intelligence for clinical knowledge retrieval: a case study using ChatGPT-4 and link retrieval plug-in to analyze diabetic ketoacidosis guidelines. Cureus 2023 Jul;15(7):e41916. [doi: 10.7759/cureus.41916] [Medline: 37457604]
13. Daei A, Soleymani MR, Ashrafi-Rizi H, Zargham-Boroujeni A, Kelishadi R. Clinical information seeking behavior of physicians: a systematic review. Int J Med Inform 2020 Jul;139:104144. [doi: 10.1016/j.ijmedinf.2020.104144] [Medline: 32334400]
14. Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. Nat New Biol 2023 Aug;620(7972):172-180. [doi: 10.1038/s41586-023-06291-2] [Medline: 37438534]

Abbreviations

AI: artificial intelligence
ED: emergency department
NHS: National Health Service
NPS: Net Promoter Score

Edited by A Kushniruk; submitted 03.09.23; peer-reviewed by C Lokker, F Alam; revised version received 14.09.24; accepted 01.12.24; published 25.02.25.

Please cite as:
Ejaz H, Tsui HLK, Patel M, Ulloa Paredes LR, Knights E, Aftab SB, Subbe CP
Comparison of a Novel Machine Learning–Based Clinical Query Platform With Traditional Guideline Searches for Hospital Emergencies: Prospective Pilot Study of User Experience and Time Efficiency
JMIR Hum Factors 2025;12:e52358
URL: https://humanfactors.jmir.org/2025/1/e52358
doi:10.2196/52358

Using a Mobile Health App (ColonClean) to Enhance the Effectiveness of Bowel Preparation: Development and Usability Study

Hui-Yu Chen[1,2*], RN, MS; Ming-Hsiang Tu[2*], RN, PhD; Miao-Yen Chen[2*], RN, PhD

[1]Endoscopy Center for Diagnosis and Treatment, Taipei Veterans General Hospital, Taipei City, Taiwan

[2]School of Nursing, National Taipei University of Nursing and Health Sciences, Room B631, No. 365, Ming-te Road, Peitou District, Taipei City, Taiwan

[*]all authors contributed equally

Corresponding Author:
Ming-Hsiang Tu, RN, PhD
School of Nursing, National Taipei University of Nursing and Health Sciences, Room B631, No. 365, Ming-te Road, Peitou District, Taipei City, Taiwan

Abstract

Background: Colonoscopy is the standard diagnostic method for colorectal cancer. Patients usually receive written and verbal instructions for bowel preparation (BP) before the procedure. Failure to understand the importance of BP can lead to inadequate BP in 25%-30% of patients. The quality of BP impacts the success of colonoscopy in diagnostic yield and adenoma detection. We developed the "ColonClean" mobile health (mHealth) app for Android devices. It incorporates visual representations of dietary guidelines, steps for using bowel cleansing agents, and observations of the last bowel movement. We used the Technology Acceptance Model to investigate whether the use of the ColonClean mHealth app can improve users' attitudes and behaviors toward BP.

Objective: This study aims to validate the effectiveness of the ColonClean app in enhancing user behavior and improving BP, providing safe and cost-effective outpatient colonoscopy guidance.

Methods: This study uses a structured questionnaire to assess perceived usefulness, perceived ease of use, and users' attitudes and behaviors toward BP regarding the ColonClean mHealth app. A total of 40 outpatients who were physically and mentally healthy and proficient in Chinese were randomly chosen for this study. The data were analyzed using SPSS 25.0, and we used Pearson product-moment correlation and simple regression analysis to predict the perception of ColonClean.

Results: The results showed that 75% (30/40) of participants achieved an "excellent" or "good" level of BP according to the Aronchick Bowel Preparation Scale. Perceived usefulness and perceived ease of use of the ColonClean mHealth app were positively correlated with users' attitudes and behaviors (P<.05).

Conclusions: The ColonClean mHealth app serves as an educational reference and enhances the effectiveness of BP. Users expressed their willingness to use the app again in the future and recommend it to family and friends, highlighting its effectiveness as an educational guide for BP.

(JMIR Hum Factors 2025;12:e58479) doi:10.2196/58479

KEYWORDS

mobile health app; bowel preparation; nursing guidance; technology acceptance model; mHealth; mobile health

Introduction

Colorectal cancer (CRC) is the third most prevalent cancer and the second leading cause of cancer-related deaths worldwide [1-3]. Colonoscopy is the standard diagnostic method for colorectal cancer, and adequate bowel preparation (BP) is a necessary and crucial step to effectively examine the entire intestinal mucosa [4,5]. Currently, BP before colonoscopy is primarily provided by nurses through paper-based nursing instructions and oral explanations of relevant precautions [6,7]. This involves dietary restrictions and the use of bowel cleansing agents to remove feces from the colon and facilitate visual examination by physicians [8,9].

The explanations for BP are often lengthy, complex, and difficult to understand and remember. As a result, the proportion of inadequate BP ranges from 25% to 30% [10]. Poor BP could lead to 42% of adenomas and 27% of advanced adenomas not being diagnosed [11,12]. Inadequate BP also leads to a missed detection of 10.55% for sessile polyps [13], which can develop into CRC and are particularly difficult to identify as they lie flat on the mucosal layer of the colon [9]. Incomplete

colonoscopy increases the risk of cancer and requires repreparation, adding to health care resource costs [14-16].

The European Society of Gastrointestinal Endoscopy guidelines emphasize that good BP is essential for the diagnosis, treatment, and removal of tumors and precancerous lesions, as well as for the reduction of CRC incidence and its mortality rate. It ensures the quality, safety, and effectiveness of colonoscopy in the clinical examination environment. Adequate BP provides optimal visualization for physicians during the examination, facilitating the detection of polyps or other lesions [10,17].

During the COVID-19 pandemic in 2020, the need for a mobile health (mHealth) app became evident [18,19]. Currently, nursing instructions for BP mainly rely on paper-based materials accompanied by oral explanations, multimedia instruction CDs, or telephone interviews [20]. Using mobile apps can enhance patient education for bowel cleansing [21,22]. Therefore, our research team conducted a literature review and assessed current needs to design and build the "ColonClean" mHealth app for the Android operating system. The purpose of this study is to validate the effectiveness of the ColonClean mHealth app in enhancing users' attitudes and behaviors toward BP through the Technology Acceptance Model (TAM) by examining the correlations among perceived usefulness, perceived ease of use, usage attitudes, and actual usage behaviors. We aim for the ColonClean mHealth app to serve as a comprehensive and effective outpatient colonoscopy guidance tool, ultimately improving the effectiveness of BP.

Methods

Ethical Considerations

This study is an interventional investigation conducted in accordance with research ethics regulations. Approval from the Institutional Review Board of Taipei Veterans General Hospital was obtained prior to the commencement of the study (approval number 2021-10-007AC). The study was conducted from October 27, 2021, to December 31, 2022. Before enrolling participants, the purpose and details of the study were explained to them, and they provided informed consent by signing a consent form. Throughout the research process, all participants' privacy rights were strictly protected, and the collected data were used solely for research purposes. Participants were not compensated for their participation in this study.

Study Participants

G*Power 3.1.9.7 for Windows was used to analyze the questionnaire data, considering a power (1-β) of 0.8, α value of .05, a medium effect size of 0.3, and an estimated dropout rate of 20%. As a result, a sample size of 40 participants was determined. A total of 40 outpatients from the gastroenterology and endoscopy center of the medical center were recruited. The inclusion criteria were patients aged 20 years or older, recommended by physicians to undergo colonoscopy, without visual or hearing impairments, and without mental illnesses. After explaining the purpose and content of the study and obtaining written informed consent, paper-based nursing instructions were provided along with verbal explanations of ColonClean, and the BP was conducted. For easy access to ColonClean, the participants downloaded the ColonClean mHealth app on their mobile devices after scanning a QR code. The nursing instruction process involved 1-on-1 teaching, with an explanation lasting approximately 20 minutes in the independent waiting area outside the outpatient visit room. Each research participant was asked to submit feedback on their experience with ColonClean within 3 days of their colonoscopy.

mHealth App (ColonClean)

Individuals aged ≥50 years are the target group for CRC screening, making them the primary users of the ColonClean mHealth app. Considering the physiological and cognitive decline in older adults, it is crucial to design a user-friendly interface to avoid difficulties in usage. The interface design for older adults incorporates intuitive elements, clear and concise text, and straightforward operational steps [23]. Contrast colors were used for interface color schemes. The operational methods are simple and clear, with functions executed through clicks or straight-line swiping gestures [24,25]. All features were designed to operate within a single page, avoiding menu structures with more than 2 layers of information. This design reduces the psychological pressure on older adults, who may be afraid of making mistakes due to an unfamiliarity with technology; this allows them to focus and read the important information comfortably. Handwritten or voice inputs were preferred for filling in the forms. Clear return or cancel buttons were placed at the top or bottom of the screen. Considering the characteristics of the users, the emphasis was on interaction and feedback rather than simple page browsing to enhance the impression of BP.

The ColonClean mHealth app was developed using the Android 11 operating system due to its low development cost and popularity among the general public [26]. The user interface and content were designed using the Android app package, the Java programming language, and a built-in database.

The screen display of ColonClean is divided into 3 sections: top, middle, and bottom in a single-page layout for the main screen. The top section displays the "Examination Date and Time." The middle section features 6 key topics arranged in a zigzag pattern based on their importance. From left to right, these topics include "Dietary Guidelines," explaining the necessary dietary education starting 3 days before the examination as shown in Figure 1, and "Laxative Use," explaining the use of bowel cleansing medication. The "Fluid Intake" screen is for recording the amount of fluid consumed (Figure 2), while the "Stool Image Recognition" screen is for providing images to identify the type of the last bowel movement (Figure 3).

Figure 1. Representative images of “Dietary Options” in ColonClean.

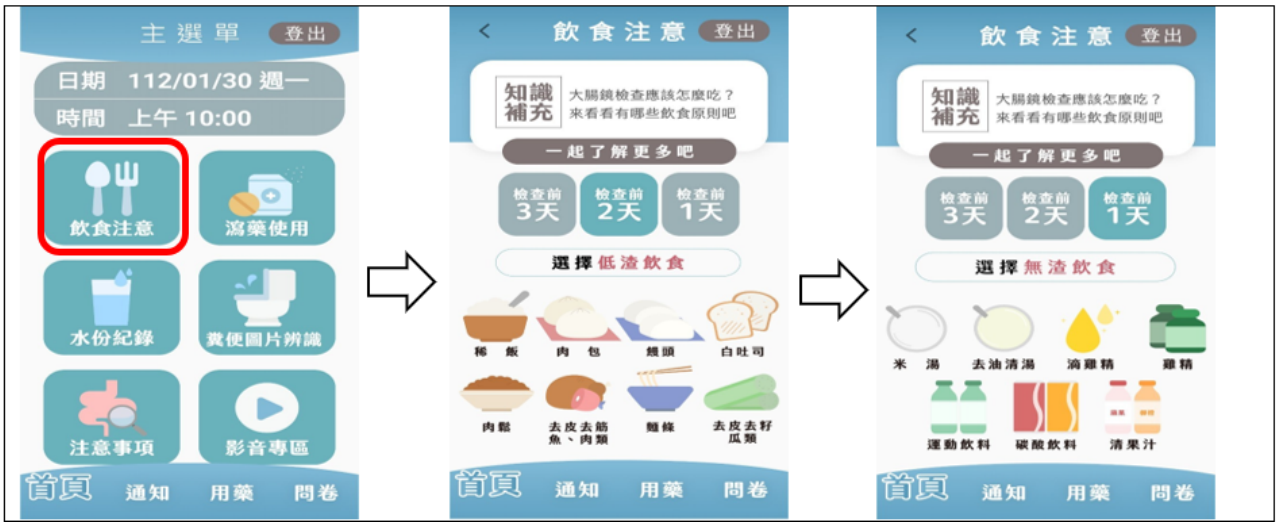


Figure 2. The “Hydration Intake Record” in ColonClean. Bowel preparation requires plenty of fluid intake to facilitate the passage of feces from the bowel.

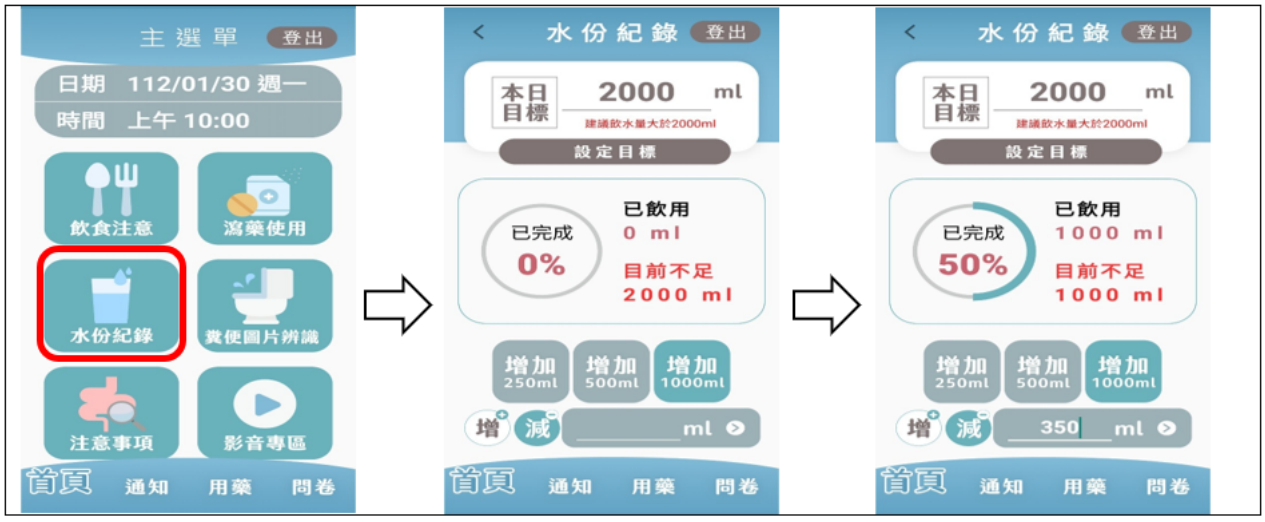
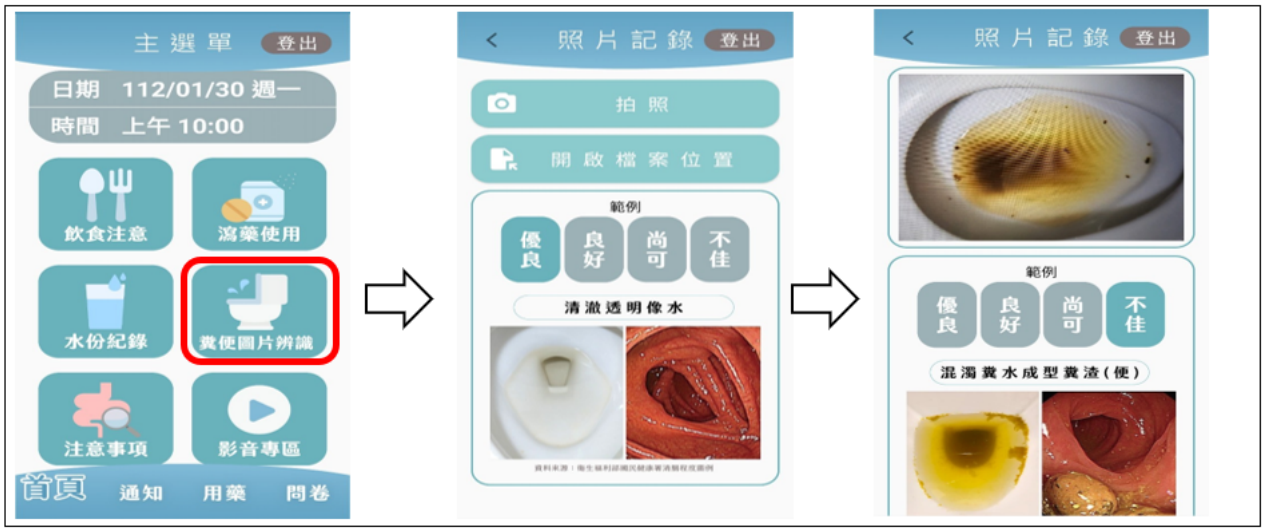


Figure 3. Images of the “Last Bowel Movement Type” section of ColonClean.



“Precautions” explains important points to keep in mind. The bottom section of the screen, from left to right, includes the “Home” button to return to the main screen; “Notifications,” which displays text messages; “Medication,” which provides

settings and modifications for regular medication or bowel cleansing agents; and "Questionnaire," which allows users to provide feedback on their experience using the ColonClean mHealth app.

The ColonClean app sent reminders and notifications to participants in the experimental group starting 3 days before their procedure. These notifications advised participants to begin dietary restrictions, such as avoiding high-fiber fruits and vegetables. Two days before the procedure, the app reminded users to switch to a low-residue diet, focusing on foods like congee or plain white noodles. The day before the procedure, the app prompted users to follow a clear liquid diet, including oil-free broth and sports drinks, and provided instructions on taking BP medications.

The app utilizes tap and swipe gestures for navigation; for user registration, it supports built-in Chinese phonetic input or handwriting recognition on mobile devices. The interface features a fresh color palette, primarily a warm lake green. Key information is highlighted in red, and unnecessary content is removed to offer concise summaries. Unique icons are designed as entry pointers to access relevant content. Tapping on an icon triggers entry into the main menu, which includes dietary guidance, instructions on using bowel cleansing agents (via images and videos), precautions for a colonoscopy, and fluid intake records, and provides references to the type of last bowel movement for understanding the differences compared to that during endoscopy. The design allows users to navigate intuitively, freely scrolling, tapping, and browsing related content between pages [27,28]. Push notifications are recommended to include instructions or prompts. Therefore, through push notifications, users are informed about relevant BP information to avoid any execution errors. Considering the user groups are middle-aged and older adults, complex actions are avoided. Instead, the key content is displayed directly in the push notification message. When a notification is received, users can click on the pushed message, which leads them to the relevant page with clear instructions or recommendations to view more detailed content. This approach reduces the chance of users misunderstanding the message. The user-centered push notification feature for BP steps enhances user engagement [29,30].

Aronchick Bowel Preparation Scale

After completing a colonoscopy, the physician submits a rating of the BP level using the Aronchick Bowel Preparation Scale (ABPS) in the endoscopy report system [1,31,32]. The ABPS is one of the most comprehensive and commonly used scales to describe the visualization of the colon and assess the percentage of stool covering the entire colon mucosa to explain the state of BP [33]. Based on this assessment, the adequacy of bowel cleansing is graded as excellent, good, fair, poor, or inadequate [11,33-35]. In Taiwan, domestic quality indicators for colonoscopy align with ABPS standards, targeting a 90% or above rating for both "excellent" and "good" categories. This ensures a high standard of care in colorectal screening, enhancing procedural accuracy and improving patient outcomes [36,37].

TAM

The TAM explains the behavior of users accepting new information technology in the field of computer technology [38]. Perceived usefulness and perceived ease of use affect users' attitudes toward adopting new technology, thereby influencing willingness and usage behavior. The perceived ease of use, perceived usefulness, attitudes, and behavior of outpatient endoscopy patients toward the ColonClean mHealth app were evaluated to assess their acceptance of using it. The questionnaire adopted a 5-point Likert scale. The 4 dimensions assessed by outpatient endoscopy patients after using ColonClean were perceived ease of use, perceived usefulness, attitudes, and usage behavior. The response options ranged from 1="strongly disagree," 2="disagree," 3="neutral," 4="agree," to 5="strongly agree." Higher scores indicated a higher level of agreement, while lower scores indicated a lower level of agreement. The questionnaire was initially developed based on literature references [39]. Three nursing informatics experts and scholars were then invited to evaluate the applicability and clarity of each question and to provide suggestions for improvement.

The Cronbach α coefficient was used to assess the internal consistency of the questionnaire items in this study. For the perceived usefulness dimension, consisting of 6 items, Cronbach α was 0.971. For the perceived ease of use dimension, consisting of 6 items, Cronbach α was 0.950. For the attitude dimension, consisting of 5 items, Cronbach α was 0.930. For the usage behavior dimension, consisting of 6 items, Cronbach α was 0.976. The questionnaire comprised a total of 23 items, and the Cronbach α coefficient for the scale was 0.986, indicating high internal consistency, meeting the requirement of a Cronbach α coefficient exceeding 0.7.

The structured questionnaire was designed and validated for reliability and validity by 3 nursing informatics experts. Participants accessed the questionnaire through a built-in link within the app, with the collected data stored in the backend management center for subsequent analysis.

Analysis

The data were analyzed using SPSS 25.0 (IBM Corp), with the utilization of the Pearson product-moment correlation and simple regression analysis to predict the perception of ColonClean.

Results

Demographic Characteristics

The demographic characteristics of the participants were as follows: 17 males (43%) and 23 females (58%); 27 participants aged ≥50 years (68%); and 35 participants with a high school/vocational education or above (88%).

ColonClean

After completing their colonoscopy, each research participant provided feedback on their experience with the ColonClean app within 3 days.. The feedback highlighted several key aspects of the app's design that contributed to user satisfaction and engagement:

- User-centered design: Participants appreciated the intuitive user interface, which maintained consistency in themes and main menu content. The clear and simple instructions, presented through a combination of text and visuals, facilitated user-friendly navigation. This consistency helped users feel more confident using the app.
- Examination date reminder: The feature displaying the examination date at the top of the main menu was well-received, as users felt it helped them stay organized and aware of the timing, potentially increasing compliance with preparation protocols.
- Graphic representations: Users found the unique graphic representations of dietary restrictions particularly helpful, as they enhanced understanding compared to traditional text descriptions. This visual approach was seen as effective in clarifying preexamination dietary options.
- Medication instructions: The clear, red-highlighted instructions for using bowel cleansing agents, along with drug images, dosage information, and timing, received positive feedback. Users appreciated the visual cues that simplified the medication process.
- Hydration tracking: Participants valued the hydration tracking feature, which encouraged them to monitor their water intake. Reminders to increase consumption when below recommended levels and positive feedback for meeting goals were highlighted as motivational elements that enhanced their commitment to the BP process.
- Bowel movement comparison: The option to take a photo of their last bowel movement and compare it with reference images was particularly beneficial. Users reported that this feature provided clarity on the required level of bowel cleanliness, enhancing their understanding of preparation requirements.
- Medication reminders and push notifications: The integration of medication reminders, particularly for chronic medications and anticoagulant discontinuation, was praised for preventing potential issues during the colonoscopy process. Users appreciated the timely push notifications that kept them informed about important tasks related to their preparation.

Overall, the research results indicated that users found the design of the ColonClean app effective in enhancing their understanding of and compliance with BP instructions. The app's focus on intuitive navigation, visual aids, and timely reminders contributed to a positive user experience, ultimately improving the effectiveness of BP for colonoscopy.

ABPS

After completing the colonoscopy, the BP level is reported by the endoscopist using the ABPS: 10 individuals (25%) achieved a "fair" level, 26 individuals (65%) achieved a "good" level, and 4 individuals (10%) achieved an "excellent" level. There were no cases of "poor" or insufficient preparation leading to interruption or inability to perform the examination.

TAM

Perceived Usefulness

The details of perceived usefulness are shown in Table 1.

Table . Perceived usefulness of the ColonClean app.

Perceived usefulness of the ColonClean app	Mean (SD)
Using ColonClean helps me quickly understand the tasks required for bowel preparation.	4.60 (0.59)
Using ColonClean provides easy access to relevant information about bowel preparation.	4.55 (0.68)
I believe that using ColonClean can improve the accuracy of my bowel preparation.	4.65 (0.53)
Using ColonClean enables me to effectively complete the required bowel preparation tasks.	4.65 (0.48)
ColonClean provides clear reminders about bowel preparation tasks, such as dietary restrictions and monitoring the type of my last bowel movement.	4.60 (0.59)
I find the educational information provided by ColonClean sufficient for completing my bowel preparation tasks.	4.65 (0.53)

The table presents users' perceptions of the usefulness of the ColonClean app in managing BP tasks. The mean scores, averaging around 4.62 of 5, suggest that users generally find the app very helpful. Below is an interpretation of the key findings:

- Understanding tasks (mean 4.60, SD 0.59): Users feel that the ColonClean app helps them quickly understand the requirements for BP, indicating clear and effective communication of instructions.
- Access to relevant information (mean 4.55, SD 0.68): The app is seen as a convenient tool for accessing important information about the preparation process. However, this metric has a slightly lower score, suggesting minor room for improvement in delivering information.
- Increased accuracy (mean 4.65, SD 0.53): Users believe that the app enhances the accuracy of their BP, reflecting a high level of trust in the app's guidance to meet medical standards.
- Effective completion of tasks (mean 4.65, SD 0.48): The app is viewed as highly effective in helping users complete BP tasks correctly, suggesting that it provides practical support throughout the process.

- Clear reminders (mean 4.60, SD 0.59): The app is viewed as highly effective in helping users complete BP tasks correctly, suggesting that it provides practical support throughout the process.
- Sufficient educational information (mean 4.65, SD 0.53): The app provides enough educational material for users to feel confident about completing their preparation, further supporting its educational role.

In summary, the ColonClean app was perceived as highly useful across various aspects of the BP process, with particularly strong scores for accuracy, task completion, and educational sufficiency. These results highlight the app's effectiveness in improving the user experience and promoting adherence to BP protocols.

Perceived Ease of Use

The details of perceived ease of use are shown in Table 2.

Table . Perceived ease of use of the ColonClean app.

Perceived ease of use of the ColonClean app	Mean (SD)
I think the interface design of ColonClean is clear and easy to understand.	4.63 (0.54)
I find the process of using ColonClean to be smooth.	4.63 (0.54)
It is easy to navigate and locate the desired functions in ColonClean.	4.63 (0.54)
I think the data presentation in ColonClean is quick and stable.	4.65 (0.58)
I find the font size in the ColonClean interface appropriate and easy to read.	4.53 (0.75)
The information provided by ColonClean makes it easier to understand the bowel preparation tasks.	4.68 (0.53)

Table 2 presents users' perceptions of the ease of use of the ColonClean app, highlighting their experiences with various interface features. The mean scores indicate a generally high level of satisfaction, with most ratings above 4.62 of 5, suggesting that users find the app intuitive and user-friendly. Below is an interpretation of the key findings:

- Clear interface (mean 4.63, SD 0.54): Users feel that the interface design is intuitive and user-friendly. This clarity in design is crucial for users' initial engagement with the app, suggesting that the developers prioritized usability.
- Smooth user experience (mean 4.63, SD 0.54): The app operates smoothly, without lag or issues, which is essential for encouraging continued use and satisfaction with digital health tools.
- Easy navigation (mean 4.63, SD 0.54): The ease of navigation suggests that users can quickly find the features they need, reducing frustration and enhancing overall satisfaction.
- Quick and stable data presentation (mean 4.65, SD 0.58): Users appreciate the app's rapid and reliable presentation of information. This reliability enhances user trust, which is particularly important in health care applications where accuracy is critical.
- Appropriate font size (mean 4.53, SD 0.75): Although the font size received a slightly lower rating, it still indicates that most users find it suitable. The slightly lower score suggests that some users might prefer larger text or different font styles, which could be an area for improvement.
- Enhanced understanding of BP (mean 4.68, SD 0.53): Users find that the information provided by ColonClean significantly aids their understanding of BP tasks, which is essential for ensuring compliance with medical procedures.

In summary, the high average scores across these metrics indicate that users generally find ColonClean to be an easy-to-use app that effectively supports BP. The highlighted areas demonstrate strengths in design, functionality, and educational support, with only minor adjustments needed in font size to further enhance the user experience. These findings suggest that ColonClean is a valuable tool for improving patient preparation and potentially contributing to better clinical outcomes.

Attitudes Toward Usage

The details of attitudes toward usage are shown in Table 3.

Table . Attitudes toward usage of the ColonClean app.

Attitudes toward usage of the ColonClean app	Mean (SD)
I believe ColonClean is an excellent choice for assisting with bowel preparation before an examination.	4.63 (0.54)
I am currently satisfied with the benefits provided by using ColonClean.	4.65 (0.53)
I would be willing to use ColonClean again in the future if I need to undergo bowel preparation.	4.63 (0.59)
I find the information provided by ColonClean to be very useful.	4.60 (0.59)
I prefer using ColonClean over paper-based nursing instructions (single sheets).	4.35 (0.77)

Table 3 summarizes users' attitudes toward the use of the ColonClean app, reflecting generally positive perceptions of its utility and effectiveness in assisting with BP tasks. The mean scores indicated a high level of satisfaction, with most ratings above 4.57 of 5, suggesting a strong endorsement of the app's features and benefits. Below is an interpretation of the key findings:

- Good choice for assistance (mean 4.63, SD 0.54): This high mean score reflects a strong consensus among users that ColonClean is an effective tool for supporting BP. The low SD suggests consistent agreement on its usefulness in this context.
- Satisfaction with benefits (mean 4.65, SD 0.53): Users' satisfaction indicates that the app meets their expectations and provides meaningful benefits during BP. The relatively low SD further shows that most users feel positively about the app's contributions.
- Willingness to use again (mean 4.63, SD 0.59): Users expressed a strong intention to use ColonClean for future BPs, suggesting that their initial positive experiences built trust in the app and that they see it as reliable.
- Usefulness of information (mean 4.60, SD 0.59): The perception that ColonClean provides highly useful information reinforces its role as a valuable educational resource, effectively aiding users in understanding the preparation process.
- Preference over paper instructions (mean 4.35, SD 0.77): Although users generally prefer ColonClean over traditional paper-based instructions, the slightly lower score and higher SD suggest that some users may still value printed materials or have mixed feelings about fully transitioning to digital formats. This variability presents an opportunity for further enhancements in the app's usability and content delivery.

In summary, the high mean scores (above 4.57) suggested that users had a positive attitude toward ColonClean, seeing it as an effective and beneficial tool for BP. The consistent responses (low SDs) indicated broad agreement among users on its usefulness. Meanwhile, the variation in preference over paper instructions may highlight areas for potential improvement. This feedback can guide future updates and enhancements to maximize user satisfaction and efficacy.

Usage Behavior

The details of usage behavior are shown in Table 4.

Table . Usage behavior of the ColonClean app.

Usage behavior of the ColonClean app	Mean (SD)
I am willing to use ColonClean as an aid for bowel preparation.	4.73 (0.45)
I am highly satisfied with the ability to access knowledge through ColonClean during bowel preparation.	4.70 (0.52)
I would recommend ColonClean to friends and family in the future.	4.63 (0.63)
I would use ColonClean to improve the effectiveness of bowel preparation.	4.70 (0.52)
I believe ColonClean provides useful information during the bowel preparation period.	4.68 (0.53)
Overall, I find using ColonClean to be satisfactory.	4.73 (0.51)

Table 4 summarizes the usage behavior of the ColonClean app and reveals overwhelmingly positive user sentiment. The overall mean score of 4.69 indicates high satisfaction with the experience of using ColonClean. This average score, combined with a relatively low SD (0.50), suggests a strong consensus among users regarding the app's effectiveness and utility. Below is an analysis of the meaning behind the scores:

- Willingness to use (mean 4.73, SD 0.45): The high mean score indicates that users are very inclined to use ColonClean as a supportive tool for BP. The low SD signifies that this sentiment is widely shared, suggesting strong confidence in the app's benefits.
- Satisfaction with knowledge access (mean 4.70, SD 0.52): Users expressed a high level of satisfaction with the knowledge accessible through the app during BP. This indicates that ColonClean effectively meets their informational needs, enhancing their preparation experience.
- Recommendation to others (mean 4.63, SD 0.63): The willingness to recommend ColonClean to friends and family reflects users' trust in the app's efficacy. Although this score remained positive, the slightly higher SD suggests some variability in users' willingness to recommend it, possibly due to differing experiences or expectations.
- Improvement of effectiveness (mean 4.70, SD 0.52): Users believed that ColonClean enhanced the effectiveness of their BP, reinforcing the app's perceived value as a helpful resource in the preparation process.
- Useful information obtained (mean 4.68, SD 0.53): The belief that the app provides useful information further underscores its utility as an educational tool. Users recognized that the app offers relevant guidance throughout the preparation period.
- Overall satisfaction (mean 4.73, SD 0.51): The high overall satisfaction score indicates that users felt positively about their experience with ColonClean. The consistency of this sentiment, reflected by the low SD, suggests that many users share a similar level of satisfaction.

In summary, the high mean scores across all categories reflect users' positive attitudes and satisfaction, underscoring ColonClean as a valuable tool for assisting with BP. The low SDs across most items indicate consistency in user experiences, suggesting that ColonClean effectively meets users' needs and

expectations. This feedback is critical for developers, as it reinforces the app's strengths and highlights areas where further enhancements could be beneficial.

Analysis

Pearson Correlation Coefficients

Pearson correlation coefficients were used to analyze the relationships among users' attitudes, perceived usefulness, perceived ease of use, and usage behavior of the ColonClean mHealth app (Table 5). The results indicated a significant positive correlation between users' attitudes and perceived usefulness (r_{38}=.907, P<.001), users' attitudes and perceived ease of use (r_{38}=.825, P<.001), users' attitudes and usage behavior (r_{38}=.835, P<.001), perceived usefulness and perceived ease of use (r_{38}=.894, P<.001), perceived usefulness and usage behavior (r_{38}=.933, P<.001), and perceived ease of use and usage behavior (r_{38}=.958, P<.001). Positive correlations were demonstrated between perceived usefulness and perceived ease of use, usage attitude, and usage behavior.

Table . Correlations among perceived usefulness, perceived ease of use, usage attitude, and usage behavior.

		Perceived usefulness	Perceived ease of use	Usage attitude	Usage behavior
Perceived usefulness					
	Pearson correlation analysis	1			
	Significance level (2-tailed)				
	Number	40			
Perceived ease of use					
	Pearson correlation analysis	0.894	1		
	Significance level (2-tailed)	<.001			
	Number	40	40		
Usage attitude					
	Pearson correlation analysis	0.907	0.825	1	
	Significance level (2-tailed)	<.001	<.001		
	Number	40	40	40	
Usage behavior					
	Pearson correlation analysis	0.933	0.958	0.835	1
	Significance level (2-tailed)	<.001	<.001	<.001	
	Number	40	40	40	40

Simple Regression Analysis

We used simple regression analysis to predict the correlation among the perceived usefulness, perceived ease of use, usage attitude, and usage behavior in the context of ColonClean. The correlation between perceived usefulness and usage attitude, with usage attitude as the criterion variable and perceived usefulness as the predictor variable, yielded B=0.907, t_{38}=13.3. This model explained 82.3% of the variance in usage attitude (R^2=0.823), and the test result with P<.001 indicated that perceived usefulness was a significant predictor variable for usage attitude. This suggests that when users perceive the app as providing substantial assistance, their attitude toward using it improves significantly. This highlights the importance of enhancing the app's practical functionality to foster a positive user experience. Similarly, the correlation between perceived usefulness and usage behavior, with usage behavior as the criterion variable and perceived usefulness as the predictor variable, resulted in B=0.933, t_{38}=16.0. This model explained 87.1% of the usage behavior (R^2=0.871), and the test result with P<.001 demonstrated that perceived usefulness significantly predicted usage behavior. When users believe that the app helps improve their preparation process, they are more likely to engage actively with it. The correlation between perceived ease of use and usage attitude, with usage attitude as the criterion variable and perceived ease of use as the predictor variable, resulted in B=0.825, t_{38}=9.0. This explained 68% of the variance in attitude toward use (R^2=0.68), with P<.001, indicating perceived ease of use as a significant predictor of attitude toward use. When users find the app easy to use, their attitude becomes more positive. This reflects their comfort and satisfaction during use, which in turn influences their willingness to engage with the

app. When determining the correlation between perceived ease of use and usage behavior, with usage behavior as the criterion variable and perceived ease of use as the predictor variable, the result was B=.958, t_{38}=20.5. This model explained 91.7% of the variance in usage behavior (R^2=.917), with P<.001, indicating that perceived ease of use significantly predicted usage behavior. When users perceive the app as easy to operate, they are more likely to engage actively. This underscores the importance of simplifying the app's interface and functionality to enhance user engagement and satisfaction.

In summary, the results indicate that both perceived usefulness and perceived ease of use have a significant positive impact on user attitudes and behaviors. This suggests that enhancing the perceived usefulness and ease of use of the ColonClean app during its design may encourage more positive user attitudes and behaviors, thereby improving health management outcomes. These findings not only validate the app's usability but also provide empirical support for future improvements.

Discussion

Principal Findings

Mobile health apps are increasingly used for illness prevention, education, and promoting healthy lifestyles [40]. However, the requirements for BP are often complex and challenging for patients to navigate and remember. To help outpatient colonoscopy patients adhere to dietary guidelines, manage bowel cleansing agents, and maintain regular medication schedules, push notifications were incorporated into the app's design. The app, protected under the patent titled "ColonClean Mobile Management System" (patent number M639094, Republic of China), features a user-centric interface that is both intuitive and visually appealing. The well-designed interface, content, push notifications, and multimedia section of the ColonClean mHealth app provide users with a convenient and helpful experience when using their mobile devices for health management. Users expressed their willingness to use the app again in the future and recommend it to family and friends, highlighting its effectiveness as an educational guide for BP.

User Attitudes and Preference for Paper-Based Instructions

Despite the overall positive reception of the ColonClean app, some users indicated a preference for paper-based instructions, as reflected in the lower mean score (mean 4.350.77) on this aspect. This finding aligns with existing literature suggesting that certain users may find physical formats more reliable or easier to reference than digital alternatives [41]. Factors contributing to this preference may include concerns about technology accessibility, device compatibility, or personal comfort with digital apps for health management.

Additionally, the preference for paper-based instructions may arise from a desire for tangible resources that can be easily annotated or highlighted [42]. Some users may feel that printed materials provide a more straightforward and immediate reference during the preparation process, particularly in situations where mobile devices are not accessible or convenient [43]. Therefore, while the digital format of the ColonClean app offers numerous advantages, such as interactivity and real-time updates, it is essential to recognize the value that traditional paper-based resources continue to hold for certain individuals.

Limitations and Recommendations

The ColonClean app operates on the Android system and functions as an mHealth app for mobile devices. However, it is important to note that during the acceptance process, instances may arise where individuals with non-Android operating systems or outdated device models are unable to be included in this study. Those unable to download ColonClean by scanning the QR code were excluded. It is advisable to consider optimizing future mHealth apps so that they can be integrated into mobile devices of various operating systems. This enhancement would contribute to more inclusive nursing guidance for BP, ensuring accessibility across platforms.

In the current era of smart health care, it is recommended to incorporate artificial intelligence recognition capabilities into the ColonClean app. Specifically, for the section where users capture images of their last bowel movement, artificial intelligence can automatically analyze and determine the quality of the stool, categorizing it as "excellent," "good," "fair," or "poor." This improvement would facilitate a more accurate assessment of the final bowel movement status. Furthermore, optimizing the ColonClean mHealth app for other individuals who require BP, such as those undergoing colorectal surgery or hospitalized patients preparing for colonoscopy, would ensure the successful completion of bowel cleansing.

Additionally, in light of the COVID-19 pandemic, when individuals undergoing screening needed to reschedule their appointments due to a positive diagnosis or being a close contact, or when they encountered issues like poor bowel movements, a consultation chat window was recommended as a future addition to address user inquiries and provide real-time solutions, thus maintaining user engagement and satisfaction. To ensure the continuity of examinations and achieve the objectives of examination, diagnosis, and treatment, it is crucial to minimize ineffective examinations or schedule cancellations that may impact individuals in need of these examinations. This approach aims to maintain a steady number of examinations and fulfill the purpose of providing necessary medical assessments and subsequent care.

Conclusions

In summary, while the ColonClean app shows promise in supporting BP, it is essential to recognize user attitudes toward traditional paper-based instructions. Enhancing the app with broader accessibility, artificial intelligence capabilities, and additional user support features could significantly improve its effectiveness and reach, ultimately contributing to better health outcomes for patients requiring BP.

Acknowledgments

We would like to express our heartfelt gratitude to the patients from the outpatient department of the gastroenterology and endoscopy center for their valuable participation in this study.

Data Availability

Data from this study are available from the corresponding author upon reasonable request.

Authors' Contributions

HYC conceptualized the study, participated in data curation and formal analysis, completed the methodology, validated the data, and drafted the manuscript. MHT supervised the study and reviewed the manuscript. MYC reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Chiu YT, Kuo CY, Lee FJ, Chang CY. Dedicated staff for patient education improves bowel preparation quality and reduces the cecal intubation time of colonoscopy: a single institution retrospective study. Medicine (Balt) 2022 Jul 22;101(29):e29437. [doi: 10.1097/MD.0000000000029437] [Medline: 35866774]
2. Mazidimoradi A, Tiznobaik A, Salehiniya H. Impact of the COVID-19 pandemic on colorectal cancer screening: a systematic review. J Gastrointest Cancer 2022 Sep;53(3):730-744. [doi: 10.1007/s12029-021-00679-x] [Medline: 34406626]
3. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021 May;71(3):209-249. [doi: 10.3322/caac.21660] [Medline: 33538338]
4. Hung SY, Chen HC, Chen WTL. A randomized trial comparing the bowel cleansing efficacy of sodium picosulfate/magnesium citrate and polyethylene glycol/bisacodyl (The Bowklean Study). Sci Rep 2020 Mar 27;10(1):5604. [doi: 10.1038/s41598-020-62120-w] [Medline: 32221332]
5. Patel N, Kashyap S, Mori A. Bowel Preparation: StatPearls; 2021. URL: https://www.ncbi.nlm.nih.gov/books/NBK535368 [accessed 2024-12-31]
6. Bechtold ML, Mir F, Puli SR, Nguyen DL. Optimizing bowel preparation for colonoscopy: a guide to enhance quality of visualization. Ann Gastroenterol 2016;29(2):137-146. [doi: 10.20524/aog.2016.0005] [Medline: 27065725]
7. Wen CC, Jao SW, Hsiao CW. A modified bowel preparation regimen for colonoscopy providing the patients' satisfaction and convenience. Med Sci Monit 2017 Jun 26;23:3123-3129. [doi: 10.12659/msm.905431] [Medline: 28649124]
8. Walter J, Francis G, Matro R, et al. The impact of diet liberalization on bowel preparation for colonoscopy. Endosc Int Open 2017 Apr;5(4):E253-E260. [doi: 10.1055/s-0043-101694] [Medline: 28382323]
9. Zad M, Do CN, Heffernan A, Johnston L, Al-Ansari M. Factors affecting bowel preparation adequacy and procedural time. JGH Open 2020 Apr;4(2):206-214. [doi: 10.1002/jgh3.12241] [Medline: 32280766]
10. Ye Z, Chen J, Xuan Z, Gao M, Yang H. Educational video improves bowel preparation in patients undergoing colonoscopy: a systematic review and meta-analysis. Ann Palliat Med 2020 May;9(3):671-680. [doi: 10.21037/apm.2020.03.33] [Medline: 32312062]
11. Sharma P, Burke CA, Johnson DA, Cash BD. The importance of colonoscopy bowel preparation for the detection of colorectal lesions and colorectal cancer prevention. Endosc Int Open 2020 May;8(5):E673-E683. [doi: 10.1055/a-1127-3144] [Medline: 32355887]
12. Sulz MC, Kröger A, Prakash M, Manser CN, Heinrich H, Misselwitz B. Meta-analysis of the effect of bowel preparation on adenoma detection: early adenomas affected stronger than advanced adenomas. PLoS One 2016;11(6):e0154149. [doi: 10.1371/journal.pone.0154149] [Medline: 27257916]
13. Chang JY, Moon CM, Lee HJ, et al. Predictive factors for missed adenoma on repeat colonoscopy in patients with suboptimal bowel preparation on initial colonoscopy: a KASID multicenter study. PLoS ONE 2018;13(4):e0195709. [doi: 10.1371/journal.pone.0195709] [Medline: 29698398]
14. Hassan C, East J, Radaelli F, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline – Update 2019. Endoscopy 2019 Aug;51(8):775-794. [doi: 10.1055/a-0959-0505] [Medline: 31295746]
15. Rex DK, Boland CR, Dominitz JA, et al. Colorectal cancer screening: recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. Gastroenterology 2017 Jul;153(1):307-323. [doi: 10.1053/j.gastro.2017.05.013] [Medline: 28600072]
16. Urban G, Tripathi P, Alkayali T, et al. Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy. Gastroenterology 2018 Oct;155(4):1069-1078. [doi: 10.1053/j.gastro.2018.06.037] [Medline: 29928897]

17. Di Leo M, Iannone A, Arena M, et al. Novel frontiers of agents for bowel cleansing for colonoscopy. World J Gastroenterol 2021 Dec 7;27(45):7748-7770. [doi: 10.3748/wjg.v27.i45.7748] [Medline: 34963739]
18. Greenhalgh T, Koh GCH, Car J. Covid-19: a remote assessment in primary care. BMJ 2020 Mar 25;368:m1182. [doi: 10.1136/bmj.m1182] [Medline: 32213507]
19. Thulesius H. Increased importance of digital medicine and eHealth during the Covid-19 pandemic. Scand J Prim Health Care 2020 Jun;38(2):105-106. [doi: 10.1080/02813432.2020.1770466] [Medline: 32484725]
20. Liu Z, Zhang MM, Li YY, Li LX, Li YQ. Enhanced education for bowel preparation before colonoscopy: a state-of-the-art review. J Dig Dis 2017 Feb;18(2):84-91. [doi: 10.1111/1751-2980.12446] [Medline: 28067463]
21. Hernández G, Gimeno-García AZ, Quintero E. Strategies to improve inadequate bowel preparation for colonoscopy. Front Med (Lausanne) 2019;6:245. [doi: 10.3389/fmed.2019.00245] [Medline: 31781565]
22. Walter B, Frank R, Ludwig L, et al. Smartphone application to reinforce education increases high-quality preparation for colorectal cancer screening colonoscopies in a randomized trial. Clin Gastroenterol Hepatol 2021 Feb;19(2):331-338. [doi: 10.1016/j.cgh.2020.03.051] [Medline: 32240835]
23. Salman HM, Wan Ahmad WF, Sulaiman S. Usability evaluation of the smartphone user interface in supporting elderly users from experts' perspective. IEEE Access 2018;6:22578-22591. [doi: 10.1109/ACCESS.2018.2827358]
24. Darvishy A, Hutter HP. Recommendations for age-appropriate mobile application design. Stud Health Technol Inform 2017;242:676-686. [Medline: 28873871]
25. Johnson J, Finn K. Designing User Interfaces for an Aging Population: Towards Universal Design: Morgan Kaufmann; 2017. [doi: 10.1016/C2015-0-01451-4]
26. Samhi J, Allix K, Bissyandé TF, Klein J. A first look at Android applications in Google Play related to COVID-19. Empir Softw Eng 2021;26(4):57. [doi: 10.1007/s10664-021-09943-x] [Medline: 33903798]
27. Bidargaddi N, Pituch T, Maaieh H, Short C, Strecher V. Predicting which type of push notification content motivates users to engage in a self-monitoring app. Prev Med Rep 2018 Sep;11:267-273. [doi: 10.1016/j.pmedr.2018.07.004] [Medline: 30109172]
28. Freyne J, Yin J, Brindal E, Hendrie GA, Berkovsky S, Noakes M. Push notifications in diet apps: influencing engagement times and tasks. Int J Hum-Comput Interact 2017 Oct 3;33(10):833-845. [doi: 10.1080/10447318.2017.1289725]
29. Back SY, Kim HG, Ahn EM, et al. Impact of patient audiovisual re-education via a smartphone on the quality of bowel preparation before colonoscopy: a single-blinded randomized study. Gastrointest Endosc 2018 Mar;87(3):789-799. [doi: 10.1016/j.gie.2017.09.007] [Medline: 28939500]
30. van der Zander QEW, Reumkens A, van de Valk B, Winkens B, Masclee AAM, de Ridder RJJ. Effects of a personalized smartphone app on bowel preparation quality: randomized controlled trial. JMIR mHealth uHealth 2021 Aug 19;9(8):e26703. [doi: 10.2196/26703] [Medline: 34420924]
31. Sharma RS, Rossos PG. A review on the quality of colonoscopy reporting. Can J Gastroenterol Hepatol 2016;2016:9423142. [doi: 10.1155/2016/9423142] [Medline: 27446877]
32. Nam SJ, Kim YJ, Keum B, et al. Impact of diet restriction on bowel preparation for colonoscopy. Medicine (Balt) 2018 Oct;97(41):e12645. [doi: 10.1097/MD.0000000000012645] [Medline: 30313052]
33. Kastenberg D, Bertiger G, Brogadir S. Bowel preparation quality scales for colonoscopy. World J Gastroenterol 2018 Jul 14;24(26):2833-2843. [doi: 10.3748/wjg.v24.i26.2833] [Medline: 30018478]
34. Manem N, Donovan K, Miller D, et al. Open-access colonoscopy quality indicators and patient perception using split-dose bowel preparation. JGH Open 2021 May;5(5):563-567. [doi: 10.1002/jgh3.12532] [Medline: 34013055]
35. Donovan K, Manem N, Miller D, et al. The impact of patient education level on split-dose colonoscopy bowel preparation for CRC prevention. J Cancer Educ 2022 Aug;37(4):1083-1088. [doi: 10.1007/s13187-020-01923-x] [Medline: 33405208]
36. Quality Indicators for Screening Colonoscopy [Website in Chinese]. Health Promotion Administration, Ministry of Health and Welfare. 2021. URL: https://www.hpa.gov.tw/File/Attach/1093/File_19179.pdf [accessed 2024-12-31]
37. May FP, Shaukat A. State of the science on quality indicators for colonoscopy and how to achieve them. Am J Gastroenterol 2020 Aug;115(8):1183-1190. [doi: 10.14309/ajg.0000000000000622] [Medline: 32759620]
38. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
39. Lai YH, Huang FF, Yang SH. Development and related research of the Chinese version of the Technology Acceptance Model scale. J Orient Inst Techn 2015;35(12):201-221 [FREE Full text]
40. Lee YL, Cui YY, Tu MH, Chen YC, Chang P. Mobile health to maintain continuity of patient-centered care for chronic kidney disease: content analysis of apps. JMIR mHealth uHealth 2018 Apr 20;6(4):e10173. [doi: 10.2196/10173] [Medline: 29678805]
41. Ferraz LT, Santos AJT, Lorenzi LJ, Frohlich DM, Barley E, Castro PC. Design considerations for the migration from paper to screen-based media in current health education for older adults: a scoping review. BMJ Open 2024 Apr 10;14(4):e078647. [doi: 10.1136/bmjopen-2023-078647] [Medline: 38604627]
42. Schorr EN, Gepner AD, Dolansky MA, et al. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the American Heart Association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103. [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]

43. Pischke CR, Voelcker-Rehage C, Ratz T, et al. Web-based versus print-based physical activity intervention for community-dwelling older adults: crossover randomized trial. JMIR mHealth uHealth 2022 Mar 23;10(3):e32212. [doi: 10.2196/32212] [Medline: 35319484]

Abbreviations

ABPS: Aronchick Bowel Preparation Scale
BP: bowel preparation
CRC: colorectal cancer
mHealth: mobile health
TAM: Technology Acceptance Model

Edited by Y Gong; submitted 18.03.24; peer-reviewed by L Weinert, YL Lee; revised version received 08.11.24; accepted 26.11.24; published 08.01.25.

Please cite as:
Chen HY, Tu MH, Chen MY
Using a Mobile Health App (ColonClean) to Enhance the Effectiveness of Bowel Preparation: Development and Usability Study
JMIR Hum Factors 2025;12:e58479
URL: https://humanfactors.jmir.org/2025/1/e58479
doi:10.2196/58479

NutriDiary, a Smartphone-Based Dietary Record App: Description and Usability Evaluation

Linda Klasen[1], MSc; Stefanie Anna Julia Koch[1], PhD; Maike Elena Benz[1], PhD; Johanna Conrad[2], PhD; Ute Alexy[1], PhD; Konrad Blaszkiewicz[2,3], MSc; Ionut Andone[2,3], MSc; Ute Nöthlings[1], DrPH

[1]Nutritional Epidemiology, Institute of Nutrition and Food Sciences, University of Bonn, Friedrich-Hirzebruch-Allee 7, Bonn, Germany

[2]German Nutrition Society (DGE), Bonn, Germany

[3]Department of Computer Science, University of Bonn, Bonn, Germany

Corresponding Author:
Ute Nöthlings, DrPH
Nutritional Epidemiology, Institute of Nutrition and Food Sciences, University of Bonn, Friedrich-Hirzebruch-Allee 7, Bonn, Germany

Abstract

Background: Repeated applications of short-term dietary assessment instruments are recommended for estimating usual dietary intake. For this purpose, NutriDiary, a smartphone app for collecting weighed dietary records (WDRs) in the German population, was developed.

Objective: We aim to describe NutriDiary and evaluate its usability and acceptability.

Methods: NutriDiary was developed as a WDR, allowing users to enter food items via text search, barcode scanning, or free text entry. The sample for the evaluation study included 74 participants (n=51, 69% female, aged 18 - 64 years), including 27 (37.5%) experts and 47 (63.5%) laypersons (including n=22, 30%, nutrition students). Participants completed a 1-day WDR and entered a predefined sample meal (n=17 foods) the following day by using NutriDiary. An evaluation questionnaire was answered from which the system usability scale (SUS) score (0 - 100) was calculated. A backward selection procedure (PROC REG in SAS; SAS Institute) was used to identify potential predictors for the SUS score (age, sex, status [expert or laypersons], and operating system [iOS or Android]).

Results: The median SUS score of 75 (IQR 63 - 88) indicated good usability. Age was the only characteristic identified as a potential predictor for a lower SUS score (P<.001). The median completion time for an individual WDR was 35 (IQR 19 - 52) minutes. Older participants took longer to enter the data than younger ones (18 - 30 y: median 1.5, IQR 1.1 - 2.0 min/item vs 45 - 64 y: median 1.8, IQR 1.3 - 2.3 min/item). Most participants expressed a preference for NutriDiary over the traditional paper-based method.

Conclusions: Good usability and acceptability make NutriDiary promising for use in epidemiological studies.

(JMIR Hum Factors 2025;12:e62776) doi:10.2196/62776

KEYWORDS

dietary assessment; food record; barcode scanning; app; mobile phone

Introduction

Usual dietary intake, the long-term average daily intake of a nutrient or food**,** is the relevant exposure when studying diet-health relationships in nutritional epidemiology. To estimate usual dietary intake, repeated applications of short-term dietary assessment instruments are recommended [1]. Self-reported dietary intake is the most commonly used method in large-scale studies through its rapid and cost-effective use. Mostly, instruments such as 24-hour dietary recalls, food frequency questionnaires or dietary records (DRs) are used, with each of these methods having individual limitations and strengths [2]. Repeated use of these traditional dietary assessment instruments is costly and burdensome for both participants and researchers [3,4]. Innovative technologies such as web- or smartphone-based tools have the potential to facilitate self-reported dietary assessment by reducing the costs and time effort of data collection and postprocessing while achieving higher acceptance in study participants [5-7]. Smartphone technology seems particularly promising among those innovative approaches. First, smartphones are widely available to a large proportion of the population. According to a survey conducted in 2021, over 95% of people aged older than 13 years in Germany already use a smartphone [8]. Second, due to the advantage of portability and the fact that most people always carry their smartphone with them, smartphone-based tools are well suited for real-time recording of food intake [1]. Third, smartphone apps can facilitate food entry through the supplementary use of an

integrated barcode scanner or by using the camera function, which reduces the need to look up every food item in a database or enter it manually [9]. Nowadays, a large number of commercial nutrition apps for self-tracking or nutritional advice are freely available [10,11]. However, because of their limited scope and questionable quality of nutrient information, those apps tend to be unsuitable for dietary assessment in epidemiological studies [12,13]. Thus, several DR apps have been developed specifically for use in epidemiological studies. In a systematic review by König et al [14], 5 core assessment features to collect data on dietary intake in scientific studies by smartphone apps were identified: photo-based assessment, assessment of serving or portion sizes, free-text description of food intake, selection from a food database, and classification systems. These features are used either alone or in combination. Thereby, the combination of photo-based recording with free-text descriptions of the consumed foods or the joint use of a food database and the assessment of serving or portion sizes were the most popular methods [14].

To provide a digital alternative for paper-based dietary assessment in epidemiological studies, we developed NutriDiary, an app for conducting weighed dietary records (WDRs) with an integrated barcode scanner. This paper aims to describe the current version of NutriDiary and to report on its usability and acceptability in laypersons and experts.

Methods

NutriDiary and the NutriDiary Database

Development and Functions of NutriDiary

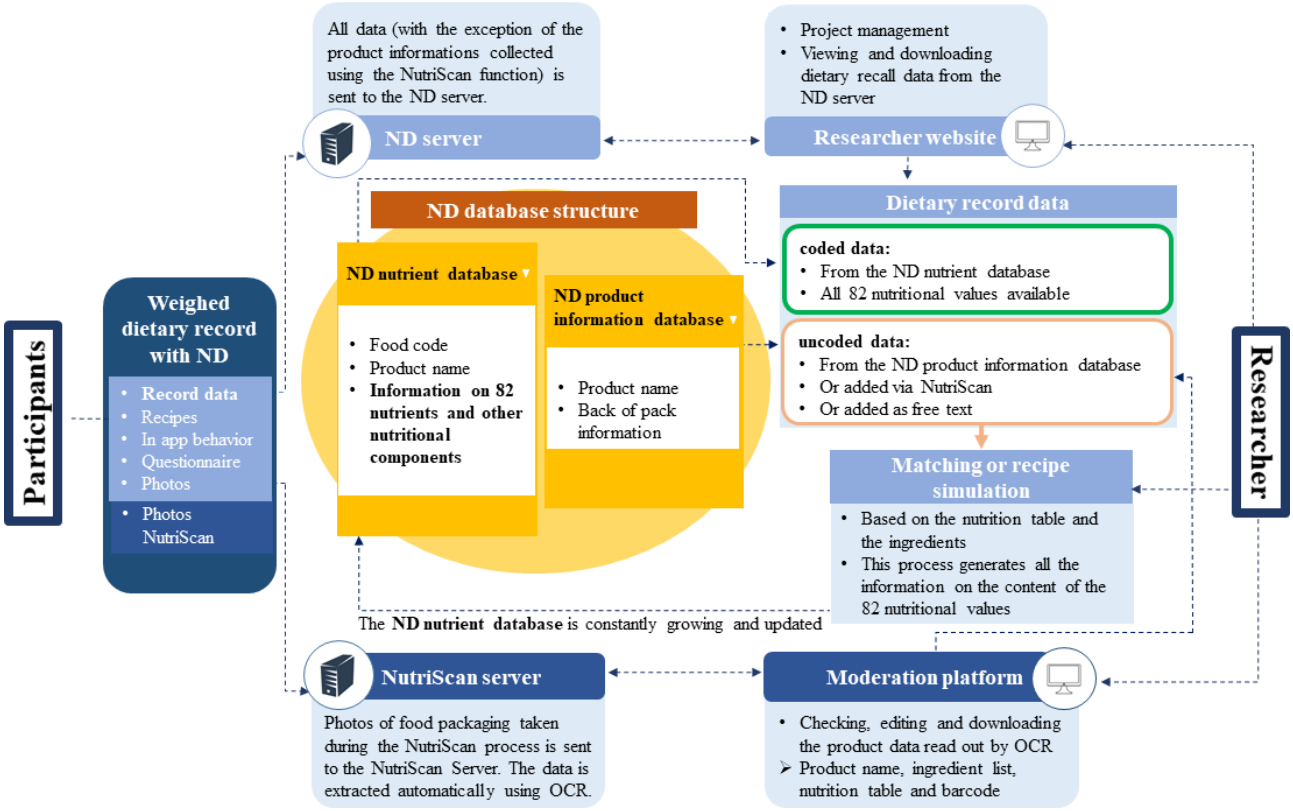
NutriDiary was developed as a smartphone app to conduct WDRs within nutritional epidemiological studies. The app is available in common app stores and study participants (hereafter referred to as users) can use the app on their smartphones with personal login data (Figure 1A). Users can start their WDR after agreeing to the data protection regulations (Figure 1B). When entering a new eating occasion, users are initially asked to enter the date, time, and place of consumption (Figure 1C). Then, the app offers three ways to enter food, beverages, and supplements: (1) a text search and subsequent selection from the underlying database, (2) barcode scanning in the underlying database, or (3) free text entry.

Figure 1. The NutriDiary app (android version): (A) screen for entering access data, (B) welcome screen, and (C) food entry.

If a food item cannot be identified via barcode scanning, the user is guided through a standardized process for collecting all relevant product information (hereinafter referred to as the NutriScan process). Thereby, the user is asked to take photos of the brand and product name, the barcode, the ingredient list and the nutrient table following step-by-step instructions (Multimedia Appendix 1). This data is then sent to the NutriScan server and automatically read out using optical character reading. Researchers can access, edit, and download this data via a moderation platform. Based on the packaging information, dieticians can match detailed nutrient data from a similar product in the database or estimate detailed nutrient values by recipe simulation in order to continuously update and expand the underlying database (Figure 2). The recipe simulation is carried out manually using the list of ingredients and the nutritional table and is described in detail elsewhere [15]. Finally, users enter the weighed amount consumed, preparation method, and quantity of potential leftovers (Figure 1C). If weighing is not possible, users are provided with a range of specified options in the drop-down menu to select an estimated portion size, such

as teaspoon or slice. After all foods and beverages have been entered, users are redirected to the main screen, where all entered eating occasions are displayed for final review. To further ease the process of recording, NutriDiary offers some usability features. A recipe editor allows entry of custom recipes, which will be added to the user's personal databases. An integrated help mode provides immediate and problem-specific assistance to users by simply touching the screen element they need help with. In addition, the app includes a photo function for collecting information, for example, on meals consumed out-of-home and entered via free text. After finishing the WDR, data are submitted to a server of the University of Bonn (NutriDiary server, Figure 2). This server also provides an administration tool to researchers (researcher website) where scientific personnel can select project-specific settings, for example, study name, study duration or the number of recording days. The app also offers the option to integrate an individual questionnaire at the end of the recording period.

Figure 2. Overview of the data structure and data flow of the NutriDiary app; ND: NutriDiary; OCR: optical character reading.



Description of the NutriDiary Database

The complex database structure of NutriDiary is shown in Figure 2. It currently contains more than 150,000 items (approximately 25% are duplicates due to different packaging sizes and different barcodes) which come from various sources. The core database is divided into the NutriDiary nutrient database and the NutriDiary product information database. The basis of the NutriDiary nutrient database is an adaptation of the in-house food and nutrient database of the DONALD (Dortmund Nutritional and Anthropometric Longitudinally Designed) study named LEBTAB (LEBensmittelTABelle, food table) [15-17]. LEBTAB currently contains around 19,000 generic and branded food items (as of June 2023) with corresponding contents of energy, 82 nutrients, and other nutritional components. Data entries for generic foods are based predominantly on the German national standard food database "Bundeslebensmittelschlüssel" (version 3.02 [18]; Federal Ministry of Food and Agriculture). The energy and nutrient contents of branded foods are predominantly estimated by recipe simulation using labeled ingredients and declared nutrient contents [15,17]. When the participant enters a food item from the NutriDiary nutrient database, this item is automatically coded and the corresponding values for energy and 82 nutrients are available. To enhance user-friendliness and allow barcode scanning, additional branded food products and barcode information were added to the NutriDiary database structure by:

- requesting product information (product name, barcode, ingredient list, and nutrition table) from food manufacturers and others (Atrify [19], a subsidiary of GS1 Germany [1WorldSync] and DATA NatuRe eG [20], a cooperative to create a central data pool for organically produced foods) to set up the NutriDiary product information database, and
- adding barcode information of already included branded products to the NutriDiary nutrient database by searching in open databases such as Open Food Facts [21], codecheck [22] (Producto Check GmbH), and Open EAN/GTIN Database (European Article Number [EAN]/Global Trade Item Number [GTIN]) [23].

For the branded products obtained through step 1 (NutriDiary product information database), information only on the nutritional values indicated on the packaging (big 7) and the ingredient list are available (Figure 2). When a participant reports one of these products in a WDR, the trained dieticians

match extended nutrient values of an equivalent food from the NutriDiary nutrient database or carry out a recipe simulation as already mentioned above [15]. This step generates all 82 nutrients for these products and they are then transferred to the NutriDiary nutrient database, which is constantly growing.

The NutriDiary Evaluation Study

Study Design

A layperson and expert evaluation was conducted. Participants were asked to keep an individual WDR with NutriDiary for 1 day and to enter a predefined sample meal on the following day. The sample meal was identical for all participants and included 4 meals (breakfast, snack, lunch, and dinner) and was provided as a digital presentation sent to study participants via email. The sample meal comprised both, generic (n=15, presented as text) and branded food items (n=3, presented as pictures of the packaging including barcode), all labeled with a hypothetical quantity of consumption, preparation method, and type of entry (text entry vs barcode scanning). Participants were generally instructed to enter branded products via scanning the barcode. Further, 1 product was intentionally not available in both NutriDiary databases, requiring participants to complete the NutriScan process in order to add the food item to the record. Furthermore, participants were tasked to correct a logged entry and to enter a hypothetical leftover without further instructions but with reference to the NutriDiary website [24] to test whether the aids provided there (frequently asked questions and help videos) fulfill their purpose (ie, helping users to help themselves). After participants completed their WDR and entered the sample meal, they were asked to answer an app-integrated evaluation questionnaire on usability and acceptability of NutriDiary on day 3. As the WDRs were not analyzed at the nutrient level, no scales were handed out for weighing the food on day 1. The participants were asked to use scales from their households.

All participants were provided with a short video giving key instructions on how to use NutriDiary before starting the WDR. The video gave a brief overview of how to use the help mode, enter foods (via text search, barcode scanning, or free text entry) and navigate the recipe book. Furthermore, participants were informed about the NutriDiary website [24], where frequently asked questions and help videos for both, iOS and Android can be found. Beyond this, participants did not undergo any additional training in using the app.

Recruitment of This Study's Population

We aimed to recruit a minimum of 51 participants, based on the assumptions outlined by Lewis and Sauro [25]. This recommendation takes into account an SD of 17.7, which is typical for system usability scale (SUS) scores [25], and ensures sufficient precision to achieve a 95% CI with a margin of error of ±5 points. This study recruited both laypersons and experts. Experts (trained nutritionists with experience in the field of dietary assessment) were recruited via direct invitations (n=28). Laypersons (n=52) were recruited via oral advertisement in lectures, mailing, and personal contact by students studying nutrition and food science at the University of Bonn, and as part of a student project. The latter mainly targeted participants between 30 and 60 years of age (in the personal environment of the students and in 2 gyms) in order to increase the number of participants in middle age and older in the group of laypersons. All participants had to be fluent in the German language, have a functioning smartphone and a valid email address. Written informed consent from all participants was obtained before enrollment.

Usability and Acceptability Assessment

The questionnaire on usability and acceptability included 14 questions on 3 different categories (usability, acceptance, and technical issues). The usability of NutriDiary was assessed by using the SUS by Brooke [26,27], which allows for comparison between similar systems or products. In short, the SUS is a Likert scale consisting of 5 positive statements (odd-numbered) and 5 negative statements (even-numbered) to which respondents indicate their degree of agreement on a scale from 1 (strong disagreement) to 5 (strong agreement). For odd items, 1 is subtracted from the user response, and for even-numbered items, 5 is subtracted from the user response, added up and multiplied by 2.5 to convert the score ranging from 0 to 100, whereas higher scores indicate better usability. According to Bangor et al [28], an SUS score below 50 is considered as "not acceptable," a score between 50 - 70 years as "marginal," and a score above 70 as "acceptable" [28,29]. In our study, an appropriate German translation of the SUS developed by SAP usability professionals of a German software corporation (SAP SE; Systems, Applications & Products in Data Processing Societas Europaea) was used and integrated into the usability questionnaire [30].

Age and sex of the participants were assessed within the evaluation questionnaire. Status (expert or layperson) was already categorized during the recruitment process. Information on the operating system was automatically recorded when NutriDiary was used and sent to the server together with the questionnaire data. To find out whether participants would prefer the app to a traditional paper-based WDR, 2 additional questions in the same structure of the SUS were added. In order to assess technical problems participants were asked whether technical errors occurred (yes or no) and, if yes, to describe the error in more detail (free text entry).

Furthermore, in-app behavior of users was recorded and evaluated. The completion time for a WDR with NutriDiary was determined by using the activity protocol of the app, in which all actions were recorded and time-stamped. For this, the time of all input activities was summed up. Interruptions of more than 5 minutes were counted as breaks and excluded from summation. As the time effort of a WDR depends on the complexity of the meals and the number of foods eaten, the relative completion time (completion time divided by the number of items) was additionally calculated. Furthermore, the percentage of estimated household measurements was of interest as well as the percentage of automatically coded WDR entries.

Statistical Analyses

Results and participants' characteristics are presented as median with their lower IQRs for continuous variables or as relative frequencies (%) for categorical variables. A backward selection

procedure (PROC REG in SAS) was used to identify characteristics of participants that were potential predictors for the SUS score. The following variables were tested: sex (male or female), age (years), status (expert or layperson), and operating system (iOS or Android). All statistical analyses were conducted using SAS (version 9.4). The significance level was set at $P<.05$.

Ethical Considerations

All examinations were carried out with written informed consent from study participants. The NutriDiary evaluation study was approved by the Ethics Committee of the University of Bonn (project identification 445/23). All data collected in this study has been pseudonymized to protect participant privacy. No reimbursement was provided to participants for their involvement in this study.

Results

From overall 80 study participants, 74 (27 experts) completed the NutriDiary evaluation study according to this study protocol and answered the evaluation questionnaire (Table 1). Most participants were female (51/74, 69%) and younger than 30 years of age (41/74, 55%).

The overall age ranged from 18 to 64 years and the median age was 29 years (IQR 25 - 45). Overall, 54% (40/74) of the participants owned a smartphone with an iOS operating system. Table 2 shows the SUS score for NutriDiary for the total study sample and stratified by sex, age groups, status, and operating system. In the overall study sample, the SUS score for NutriDiary ranged from 43 to 100 (data not shown). The median SUS score of 75 (IQR 63 - 88) indicates a good usability of NutriDiary. The mean SUS score was 74 (SD 15; 95% CI 70-77; data not shown). The median SUS score was higher in women than in men (80, IQR 65 - 88, vs 70, IQR 55 - 78) and higher in the group of experts than in the group of laypersons (80, IQR 65 - 88 vs 73, IQR 58 - 88), with female laypersons having a higher SUS score than female experts (83, IQR 66 - 88 vs 80, IQR 60 - 85). Looking at the group of laypersons after excluding nutrition students (labeled in Tables 1 and 2 as “others,” median age 55, IQR 29 - 58, years), the median SUS score was 63 (IQR 50 - 73).

Table . Characteristics of study participants of the NutriDiary evaluation study (N=74).

	Total	Experts	Laypersons		
			Total	Nutrition students	Others
Participants, n (%)	74 (100)	27 (36.5)	47 (63.5)	22 (47)	25 (53)
Sex (group), n (%)					
Women	51 (69)	23 (85)	28 (60)	19 (86)	9 (36)
Age, median (IQR)	29 (25 - 45)	35 (28 - 40)	27 (23 - 56)	23 (22 - 26)	55 (29 - 58)
Age (years), n (%)					
Age group 1: 18-≤30	41 (55.4)	12 (44.4)	29 (61.7)	22 (100)	7 (28)
Age group 2:>30-≤45	15 (20.3)	11 (40.7)	4 (8.5)	0 (0)	4 (16)
Age group 3:>45 - 65	18 (24.3)	4 (14.8)	14 (29.8)	0 (0)	14 (56)
Operating system, n (%)					
iOS	40 (54)	12 (44.4)	28 (59.6)	12 (54.5)	16 (64)
Android	34 (46)	15 (55.6)	19 (40.4)	10 (45.5)	9 (36)

Table . System usability scale (SUS) score for the NutriDiary app presented as median (IQR) values.

	Total		Women		Men	
	n	Median (IQR)	n	Median (IQR)	n	Median (IQR)
SUS score	74	75 (63 - 88)	51	80 (65 - 88)	23	70 (55 - 78)
Stratified by age group						
Age group 1: 18 -≤30 years	41	80 (70 - 88)	31	83 (70 - 88)	10	71 (63 - 85)
Age group 2: >30 -≤45 years	15	80 (68 - 90)	11	85 (65 - 93)	4	74 (70 - 78)
Age group 3: >45 - 65 years	18	58 (50 - 70)	9	58 (48 - 65)	9	65 (50 - 70)
Stratified by status						
Experts	27	80 (65 - 88)	23	80 (60 - 85)	4	83 (70 - 90)
Laypersons (total)	47	73 (58 - 88)	28	83 (66 - 88)	19	70 (55 - 73)
Nutrition students	22	85 (75 - 88)	19	88 (78 - 90)	3	73 (70 - 85)
Others	25	63 (50 - 73)	9	55 (48 - 65)	16	66 (53 - 73)
Stratified by operating system						
iOS	40	78 (65 - 88)	26	80 (68 - 90)	14	66 (55 - 73)
Android	34	74 (58 - 85)	25	75 (58 - 85)	9	73 (70 - 85)

Among the age groups 1 (18-≤30 y) and 2 (>30-≤45 y), the median SUS score was noticeably higher (80, IQR 70 - 88 and 80, IQR 68 - 90) than in age group 3 (58, IQR 50 - 70). The results of the backward selection procedure showed that age was the only characteristic identified as a potential predictor for the SUS score in the examined sample (P<.001).

Figure 3 shows the individual statements of the SUS questionnaire and a summary of the answers given by the 74 participants, presented as box plots. The figure shows that agreement tends to be high for the positive (odd-numbered) statements and tends to be low for the negative (even-numbered) statements, as is a prerequisite for a higher SUS score. Among the positive statements, the participants showed the lowest level of agreement with statement 1 “I think that I would like to use NutriDiary frequently.” For the negative statements, agreement was highest for statement 8 “I found NutriDiary very cumbersome to use.”

Figure 3. User rating (N=74) of the individual statements of the SUS evaluating the usability and acceptability of NutriDiary, shown as box plots. The dot represents the mean and the thick stripe the median. SUS: system usability scale.

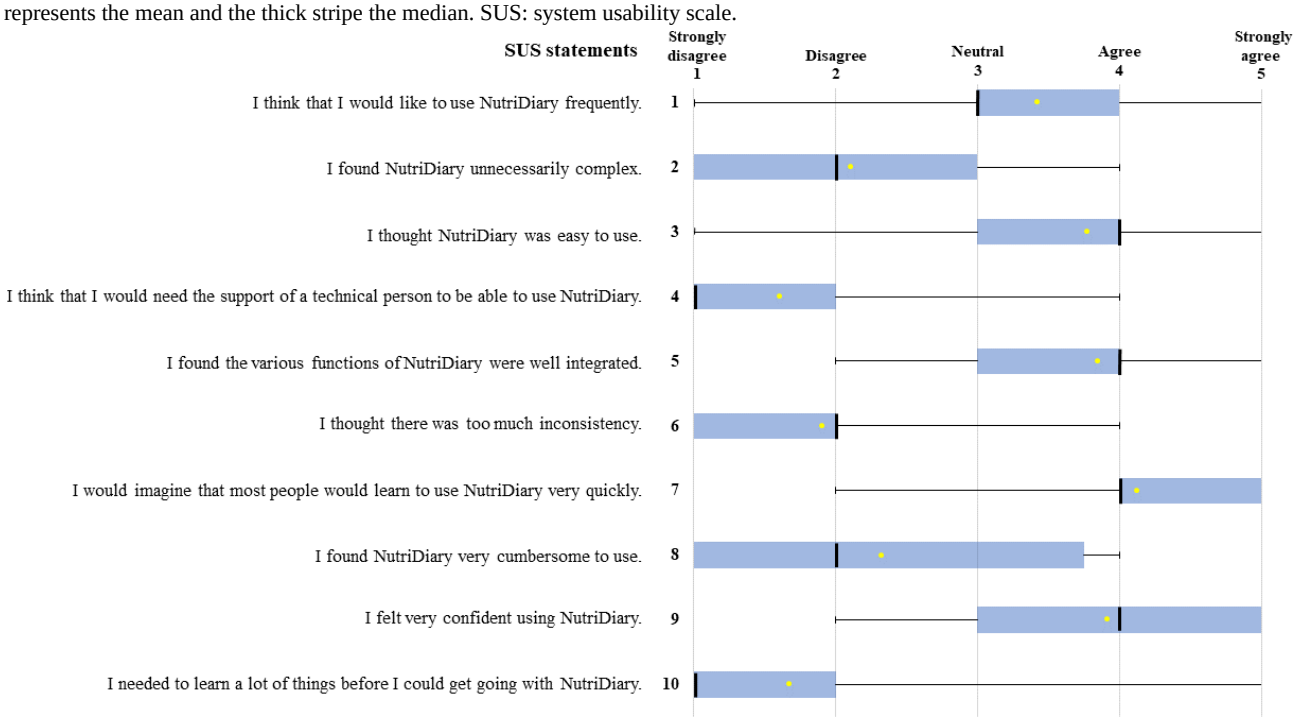


Table 3 shows the results of the evaluated in-app behavior logs and WDRs. The calculated median completion time for an individual WDR (day 1) with NutriDiary was 35 (IQR 19 - 52) minutes. Participants needed a median of 1.6 (IQR 1.2 - 2)

minutes to enter an item. This relative completion time was slightly higher in age group 3 (>45 - 63 y) compared to younger participants (1.8, IQR 1.3 - 2.3, min/item vs 1.5, IQR 1.1 - 2, min/item). The median time it took participants to enter the sample meal in NutriDiary on day 2 was 15 (IQR 12 - 18) minutes. The median relative completion time for entering the sample meal was 0.8 (IQR 0.7 - 0.9) min/item. Participants in age group 3 took slightly longer than younger participants. The median proportion of already coded items was 68% (IQR 55% - 82%) and the median proportion of estimated quantities (not weighed) was 8% (IQR 0% - 29%). Younger participants (age group 1) used more estimated household measurements instead of weighing than participants in age groups 2 and 3 (0% - 26%, 8% vs 0% - 31%, 7% and 0% - 23%, and 4%, respectively).

Table . Completion time and proportion of items already coded for the total sample (N=74) and stratified by age groups.

	Total	Age group 1: 18 - 30 y	Age group 2: 31 - 45 y	Age group 3: 46 - 65 y
Participants, n (%)	74 (100)	41 (55.4)	15 (20.3)	18 (24.3)
Number of entered items day 1[a]	21 (15 - 31)	22 (16 - 31)	21 (19 - 32)	21 (13 - 30)
Total completion time day 1 (min)[a]	35 (19 - 52)	35 (19 - 45)	44 (19 - 55)	36 (17 - 53)
Relative completion time day 1 (min/item)[a]	1.6 (1.2 - 2)	1.5 (1.1 - 2)	1.5 (1.1 - 2)	1.8 (1.3 - 2.3)
Total completion time sample meal (min)[a]	15 (12 - 18)	14 (11 - 17)	17 (11 - 19)	17 (13 - 23)
Relative completion time sample meal (min/item)[a]	0.8 (0.7 - 0.9)	0.8 (0.6 - 0.9)	0.9 (0.6 - 1.1)	0.9 (0.7 - 1.3)
Proportion of items already coded (%)[a]	68 (55 - 82)	63 (56 - 79)	77 (62 - 86)	71 (50 - 86)
Proportion of estimated household quantities (%)[a]	8 (0 - 29)	8 (0 - 26)	7 (0 - 31)	4 (0 - 23)

[a]Median (IQR).

In total, 24% (18/74) of the participants reported technical issues. In 2 cases, the app crashed, but could be reopened afterward and no data was lost. In 5 cases, participants stated that they had problems finding newly added products (this process is sometimes delayed in rare cases). Further, 4 participants reported that the selection from the drop-down menu did not work properly. In the remaining cases, participants described issues rather associated with this study design than with the technical functions of NutriDiary or gave general comments. For example, 4 participants reported that a barcode presented in the sample meal could not be scanned from the screen (probably due to varying screen-brightness and screen-resolution).

Overall, 85% (63/74) of the participants were able to enter the sample meal correctly. Of the 11 participants who made mistakes when entering the sample meal, 5 were in age group 1, 2 in age group 2, and 4 in age group 3, indicating that age did not affect the accuracy of data entry. In 2 cases, food items were missing. Further, 3 participants had difficulties in editing a WDR entry and entering hypothetical leftovers. Furthermore, 6 participants entered a different amount of food than that presented in the sample meal. When participants were asked whether they would prefer NutriDiary to the traditional paper method, a total of 77% (57/74) agreed. Only 9.5% (7/74) could imagine, that it would be easier to keep a WDR with pen-and-paper than to use NutriDiary (Multimedia Appendix 2).

Discussion

Principal Findings

To our knowledge, an app-based WDR system with barcode scanning function for scientific studies in Germany does not exist so far. The popularity and widespread use of smartphones make NutriDiary a promising alternative to the traditional pen-and-paper approach. Different ways for entering food items (text search, barcode scanning, and free text entry) and the NutriScan function enable users to record the products they consume in a very detailed manner. The digital data output has the potential to reduce the burden for researchers also. In this study, we evaluated NutriDiary in a convenience sample of experts and laypersons and found good usability and acceptability.

The median SUS score of 75 indicated good usability of NutriDiary. The SUS score has been developed as a means to measure the overall perceived usability of a system [26,27]. However, technology-based dietary assessment instruments for scientific purposes are very specific tools that are not designed for everyday use but to generate scientifically useful data. The collection of these data is often challenging for the participants. The first statement of the SUS questionnaire is “I think that I would like to use this system frequently.” This statement is unlikely to be agreed upon by many people in the case of a WDR due to the burden on participants by the method itself.

Nevertheless, the advantage of the SUS score is that it offers the possibility to compare systems used in the same context [27]. The median SUS score of 75 for NutriDiary (mean SUS score: 74, SD 15, for comparison) is comparable with SUS scores for other similar technology-based food records such as "METADIETA-web" (mean SUS score: 68, SD 15, n=26) [31], the "Eat and Track app" (mean SUS score: 69, n=15) [32] or the "Traqq app" (mean SUS score: 79, SD 15, n=22) [33].

Age has been discussed as a major limiting factor for the use of technology-based systems [29] or innovative dietary assessment instruments [31,34,35]. Consistently, the results of the backward selection procedure showed that age was identified as a potential predictor for a lower SUS score in the examined sample. Further, older participants took slightly longer to complete a WDR with NutriDiary than younger ones. Feasibility testing of the digital food record METADIETA-web by Vitale et al [31] also showed that the preference for using the digital tool instead of the traditional pen-and-paper method decreased with increasing age [31]. Older generations did not grow up with smartphones and computers and are therefore likely to be less intuitive with apps and web-based tools than younger generations. However, longer completion times could also be explained by higher accuracy in entering items and greater patience [36], which would be in line with the observation that older participants used fewer estimated household measurements instead of weighing than younger ones in our study (0% - 23%, 4% for age group 3 vs 0% - 31%, 7% and 0% - 26%, 8% for age group 2 and 1, respectively). Nevertheless, an understandable introduction to the use of the instrument might be particularly important in the group of older adults. For practical reasons, only an introductory video was sent out in this study. Our experience in training users to use the NutriDiary app shows that older people tend to ask more questions. This can only be addressed in a face-to-face conversation where it can be ensured that all relevant information is conveyed. Therefore, we recommend a personal introduction when using NutriDiary in epidemiological studies, if feasible.

The median SUS score was higher in women than in men and higher in the group of experts than in the group of laypersons. This may partly be explained by the fact that the proportion of older people was higher in men than in women and higher in the group of laypersons than in the group of experts. The group of laypersons also included students of nutritional science. Although this group certainly cannot be described as experts, the students may have some background knowledge that could have influenced the outcome. Whether this makes them more critical or less critical in their judgment remains questionable.

WDRs have the potential to provide the most accurate description of the types and amounts of the foods consumed over a specified period of time, but they are also considered to be one of the most burdensome and elaborate dietary assessment methods. When conducting a WDR, participants have to weigh and write down everything they eat and drink and always carry their kitchen scales, the record sheet, and a pen with them. This process is exhausting and time-consuming and requires a high level of cooperation from the participants [37]. The median completion time for conducting a WDR with NutriDiary was 35 minutes (1.6 min/item, Table 3) and can be rated as acceptable. When looking at the median relative completion time of the sample meal (here, the amount of the presented foods was predefined), it can be assumed that the single entering process (without weighing) takes around 0.8 minutes per item. This result suggests that weighing accounts for about half the completion time. Keeping this in mind, the single entering process is roughly comparable with the reported average completion time of other text-based DR apps, for example, with "My Meal Mate." Here, participants needed on average 22 minutes per day to complete a record with estimated (not weighed) portion sizes [38].

Whether the NutriDiary app significantly reduces the completion time compared to the traditional pen-and-paper method is questionable, because the digital version of a WDR does not change the fact that weighing is required for this dietary assessment method. However, NutriDiary offers some advantages that can make its use more attractive than the traditional method. First, most people always carry their smartphone with them and recording "on the go" is more practical than using pen-and-paper. Presumably, this makes it less likely that participants forget their records and need to add the meal at a later time. Second, features such as the recipe book and the barcode scanner can make recording easier, as frequently consumed food combinations and recipes can be stored and retrieved later, and consumed products can be added more easily. Furthermore, integrated standard household measurements make it easier to estimate the quantity if weighing is not possible, for example, when eating out of home. The photo function allows participants to add a photo of the meal, which can help the postprocessing. To figure out whether NutriDiary is more attractive to study participants than the traditional pen-and-paper method, we added 2 more questions to the SUS questionnaire (, Multimedia Appendix 2). The result clearly showed that the vast majority would prefer using the NutriDiary app instead of a paper-based food record. To avoid errors caused by application problems using technology-based instruments, an understandable introduction and information structure are necessary to prevent frustration and enable participants to help themselves quickly if they have difficulties. For this purpose, we designed a website where all information and support materials are centralized in 1 location [24].

When using traditional DRs in epidemiological studies, the postprocessing of the DRs is very time-consuming. The data needs to be digitized and manually coded by the study staff. In this study, 68% (IQR 55% - 82%) of the food entered in NutriDiary was already coded. Furthermore, the data was already available in digital form, suggesting less costs and time in data postprocessing, compared to the pen-and-paper method. NutriDiary was designed as an app for keeping WDRs, meaning that participants are asked to weigh all the food and drinks they consume. However, as experience showed that this is not always possible, the app also offers the choice of standard household measurements (eg, teaspoon, glass, or portion) to estimate the quantity in situations where weighing is not possible. Considering the burden of weighing all foods, this selection option might tempt participants to estimate rather than weigh. However, the proportion of estimated quantities was rather low and can be considered manageable in this study.

Within the development process of NutriDiary, building an appropriate food database structure and the collection, standardization, and integration of barcode and food packaging information to enable food entry via barcode scanning was one of the most challenging tasks. The underlying database determines the users' success in searching for food items and is essential for the functionality and accuracy of technology based, self-administered dietary assessment tools [3,39]. Therefore, we aimed to make the database as complete as possible. Due to a high number of available branded foods and the frequently changing food market, this is a major challenge [12,40-43]. According to the Food Federation Germany [44], there are more than 170,000 food products available on the German food market. Every year, about 40,000 new products are launched and just as many disappear [44]. It also happens that manufacturers change food recipes, which can also change the ingredient lists and nutritional values. These circumstances mean that the database needs to be constantly updated. To make this possible, the NutriScan function described above was developed and integrated. If the app is used regularly, the database is regularly updated with new products recorded by users.

Strengths and Limitations

The NutriDiary evaluation study had some strengths and limitations. As NutriDiary was developed for scientific use, practicing nutritionists were also recruited as experts. Although experts are not intended users of the app, their evaluation allows for a professional view of the usability of NutriDiary. Furthermore, experts decide on the use of the assessment instrument in scientific studies, which is why their opinion on usability is of particular importance in this context. To assess user-friendliness in older age groups as well, we specifically recruited older participants. However, this study population was not representative of the general German population. Most participants were female, young, and highly educated. This limited the generalizability of our results. Nevertheless, it is important to acknowledge that participants in epidemiological studies frequently possess a higher educational status compared to the general population. In addition, participants were aware of the purpose of this study, which was to evaluate the usability and acceptability of NutriDiary, making it difficult to assess the quality and transferability of the data on the completion time and proportion of items already coded. This study aimed to assess the usability and acceptability of NutriDiary and does not contain any data on the validity of the NutriDiary app. However, providing information on the app's validity and psychometric properties is a critical prerequisite for its effective application. Therefore, a validation study has already been initiated and will give insights into the validity and quality of the assessed nutritional data.

Conclusion

NutriDiary is the first smartphone based WDR app with integrated barcode scanning function for scientific purposes in Germany. The evaluation by experts and laypersons indicated an acceptable completion time, good usability and acceptability on the users' side, whereby younger experts and laypersons tended to rate the app better than older ones. Future research will give insights into the validity and feasibility of NutriDiary in different study populations.

Acknowledgments

The authors would like to thank all study participants for the provision of their data. Special thanks go to the students Sofie Frey, BSc, and Franziska Oidtmann, BSc, for their support with data collection. This work was supported by Diet-Body-Brain (DietBB) and the Competence Cluster in Nutrition Research funded by the Federal Ministry of Education and Research (FKZ: 01EA1410A).

Authors' Contributions

LK and UN did the conceptualization. IA, KB, JC, SAJK, LK, MEB, and UN developed the app. LK and MEB handled data curation and worked on the formal analyses. LK wrote the original draft. SAJK, MEB, UA, JC, IA, and UN reviewed and edited the writing. UN supervised the study.

Conflicts of Interest

The authors were involved in the development of the NutriDiary app.

Multimedia Appendix 1
The NutriScan process.
[PNG File, 73 KB - humanfactors_v12i1e62776_app1.png]

Multimedia Appendix 2
User rating (n=74) on participants' preferences for NutriDiary compared to a traditional WDR (pen -and -paper method).
[JPEG File, 90 KB - humanfactors_v12i1e62776_app2.jpeg]

References

1. Conrad J, Nöthlings U. Innovative approaches to estimate individual usual dietary intake in large-scale epidemiological studies. Proc Nutr Soc 2017 Aug;76(3):213-219. [doi: 10.1017/S0029665116003025] [Medline: 28162110]

2. Cade JE. Measuring diet in the 21st century: use of new technologies. Proc Nutr Soc 2017 Aug;76(3):276-282. [doi: 10.1017/S0029665116002883] [Medline: 27976605]
3. Koch SAJ, Conrad J, Hierath L, et al. Adaptation and evaluation of Myfood24-Germany: a web-based self-administered 24-h dietary recall for the German adult population. Nutrients 2020 Jan 6;12(1):160. [doi: 10.3390/nu12010160] [Medline: 31935885]
4. Park Y, Dodd KW, Kipnis V, et al. Comparison of self-reported dietary intakes from the automated self-administered 24-h recall, 4-d food records, and food-frequency questionnaires against recovery biomarkers. Am J Clin Nutr 2018 Jan 1;107(1):80-93. [doi: 10.1093/ajcn/nqx002] [Medline: 29381789]
5. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. Nutrition 2014;30(11-12):1257-1266. [doi: 10.1016/j.nut.2014.02.020] [Medline: 24976425]
6. Illner AK, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. Int J Epidemiol 2012 Aug;41(4):1187-1203. [doi: 10.1093/ije/dys105] [Medline: 22933652]
7. Thompson FE, Subar AF, Loria CM, Reedy JL, Baranowski T. Need for technological innovation in dietary assessment. J Am Diet Assoc 2010 Jan;110(1):48-51. [doi: 10.1016/j.jada.2009.10.008] [Medline: 20102826]
8. Tenzer F. Smartphone-Besitz bei Kinder und Jugendliche in Deutschland im Jahr 2021 nach Altersgruppe. Statista. 2022. URL: https://de.statista.com/statistik/daten/studie/1106/umfrage/handybesitz-bei-jugendlichen-nach-altersgruppen/ [accessed 2025-01-22]
9. Mortazavi BJ, Gutierrez-Osuna R. A review of digital innovations for diet monitoring and precision nutrition. J Diabetes Sci Technol 2023 Jan;17(1):217-223. [doi: 10.1177/19322968211041356] [Medline: 34467803]
10. Holzmann SL, Pröll K, Hauner H, Holzapfel C. Nutrition apps: quality and limitations. An explorative investigation on the basis of selected apps. Ernahrungs Umschau 2017;64(5):80-89. [doi: 10.4455/eu.2017.018]
11. Ferrara G, Kim J, Lin S, Hua J, Seto E. A focused review of smartphone diet-tracking apps: usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. JMIR mHealth uHealth 2019 May 17;7(5):e9232. [doi: 10.2196/mhealth.9232] [Medline: 31102369]
12. Maringer M, Wisse-Voorwinden N, Geelen A. Food identification by barcode scanning in the Netherlands: a quality assessment of labelled food product databases underlying popular nutrition applications. Public Health Nutr 2019 May;22(7):1215-1222. [doi: 10.1017/S136898001800157X] [Medline: 29962361]
13. Fallaize R, Zenun Franco R, Pasang J, Hwang F, Lovegrove JA. Popular nutrition-related mobile apps: an agreement assessment against a UK reference method. JMIR mHealth uHealth 2019 Feb 20;7(2):e9838. [doi: 10.2196/mhealth.9838] [Medline: 30785409]
14. König LM, Van Emmenis M, Nurmi J, Kassavou A, Sutton S. Characteristics of smartphone-based dietary assessment tools: a systematic review. Health Psychol Rev 2022 Dec;16(4):526-550. [doi: 10.1080/17437199.2021.2016066] [Medline: 34875978]
15. Sichert-Hellert W, Kersting M, Chahda C, Schäfer R, Kroke A. German food composition database for dietary evaluations in children and adolescents. J Food Compost Anal 2007 Feb;20(1):63-70. [doi: 10.1016/j.jfca.2006.05.004]
16. Kroke A, Manz F, Kersting M, et al. The DONALD Study. History, current status and future perspectives. Eur J Nutr 2004 Feb;43(1):45-54. [doi: 10.1007/s00394-004-0445-7] [Medline: 14991269]
17. Perrar I, Alexy U, Nöthlings U. Cohort profile update-overview of over 35 years of research in the Dortmund Nutritional and Anthropometric Longitudinally Designed (DONALD) study. Eur J Nutr 2024 Apr;63(3):727-740. [doi: 10.1007/s00394-023-03290-x] [Medline: 38151532]
18. Bundeslebensmittelschnlüssel (BLS), version 3.02. Federal Ministry of Food and Agriculture. URL: https://www.blsdb.de/ [accessed 2025-01-22]
19. GDSN plattform für GS1 member organisationen und communities. 1WorldSync. 2023. URL: https://www.atrify.com/partner/gs1-member-organizations/?gclid=CjwKCAjwqZSlBhBwEiwAfoZUIJ8s4-oKxmTJ2Q86jycpue1mf7O2__wWg6zVBsrmh0KQ_kBbV0gloRoCPXQQAvD_BwE [accessed 2025-01-22]
20. DATA NatuRe eG. 2023. URL: https://datanature.de/ [accessed 2025-01-22]
21. Open food facts. 2023. URL: https://de.openfoodfacts.org/discover [accessed 2025-01-22]
22. Codecheck. 2023. URL: https://www.codecheck.info/ [accessed 2025-01-22]
23. Open EAN/GTIN Database. 2023. URL: https://opengtindb.org/ [accessed 2025-01-22]
24. Universität Bonn. NutriDiary. 2024. URL: https://www.nutridiary-app.de [accessed 2025-01-22]
25. Lewis J, Sauro J. Sample sizes for a SUS score. MeasuringU. 2022. URL: https://measuringu.com/sample-sizes-for-sus-ci [accessed 2025-01-22]
26. Brooke J. SUS-a quick and dirty usability scale. URL: https://hell.meiert.org/core/pdf/sus.pdf [accessed 2025-01-22]
27. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text] [doi: 10.5555/2817912.2817913]
28. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J User Exper 2009;4(3):114-123 [FREE Full text] [doi: 10.5555/2835587.2835589]

29. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 29;24(6):574-594. [doi: 10.1080/10447310802205776]
30. System usability scale – jetzt auch auf deutsch. SAP Website. 2023. URL: https://blogs.sap.com/2016/02/01/system-usability-scale-jetzt-auch-auf-deutsch/ [accessed 2025-01-22]
31. Vitale M, Bruno V, D'Abbronzo G, et al. Evaluation of eating habits by 7-day food record: web-PC vs. traditional paper format. Int J Food Sci Nutr 2023;74(4):580-587. [doi: 10.1080/09637486.2023.2236808] [Medline: 37455370]
32. Wellard-Cole L, Potter M, Jung JJ, Chen J, Kay J, Allman-Farinelli M. A tool to measure young adults' food intake: design and development of an australian database of foods for the eat and track smartphone app. JMIR mHealth uHealth 2018 Nov 7;6(11):e12136. [doi: 10.2196/12136] [Medline: 30404768]
33. Lucassen DA, Brouwer-Brolsma EM, van de Wiel AM, Siebelink E, Feskens EJM. Iterative development of an innovative smartphone-based dietary assessment tool: Traqq. J Vis Exp 2021 Mar 19(169). [doi: 10.3791/62032] [Medline: 33818566]
34. Gilsing A, Mayhew AJ, Payette H, et al. Validity and reliability of a short diet questionnaire to estimate dietary intake in older adults in a subsample of the canadian longitudinal study on aging. Nutrients 2018 Oct 17;10(10):1522. [doi: 10.3390/nu10101522] [Medline: 30336568]
35. Carter MC, Albar SA, Morris MA, et al. Development of a UK online 24-h dietary assessment tool: myfood24. Nutrients 2015 May 27;7(6):4016-4032. [doi: 10.3390/nu7064016] [Medline: 26024292]
36. Kirkpatrick SI, Gilsing AM, Hobin E, et al. Lessons from studies to evaluate an online 24-hour recall for use with children and adults in canada. Nutrients 2017 Jan 31;9(2):100. [doi: 10.3390/nu9020100] [Medline: 28146125]
37. Rutishauser IHE. Dietary intake measurements. Public Health Nutr 2005 Oct;8(7A):1100-1107. [doi: 10.1079/phn2005798] [Medline: 16277822]
38. Carter MC, Burley VJ, Nykjaer C, Cade JE. "My Meal Mate" (MMM): validation of the diet measures captured on a smartphone application to facilitate weight loss. Br J Nutr 2013 Feb 14;109(3):539-546. [doi: 10.1017/S0007114512001353] [Medline: 22717334]
39. Kupis J, Johnson S, Hallihan G, Olstad DL. Assessing the usability of the Automated Self-Administered Dietary Assessment Tool (ASA24) among low-income adults. Nutrients 2019 Jan 10;11(1):132. [doi: 10.3390/nu11010132] [Medline: 30634581]
40. Ocké MC, Westenbrink S, van Rossum CTM, Temme EHM, van der Vossen-Wijmenga W, Verkaik-Kloosterman J. The essential role of food composition databases for public health nutrition – experiences from the Netherlands. J Food Compost Anal 2021 Aug;101:103967. [doi: 10.1016/j.jfca.2021.103967]
41. Westenbrink S, van der Vossen-Wijmenga W, Toxopeus I, Milder I, Ocké M. LEDA, the branded food database in the Netherlands: data challenges and opportunities. J Food Compost Anal 2021 Sep;102:104044. [doi: 10.1016/j.jfca.2021.104044]
42. Pravst I, Hribar M, Žmitek K, Blažica B, Seljak BK, Kušar A. Branded foods databases as a tool to support nutrition research and monitoring of the food supply: insights from the Slovenian composition and labeling information system. Front Nutr 2021;8:798576. [doi: 10.3389/fnut.2021.798576] [Medline: 35059426]
43. Cade JE, Warthon-Medina M, Albar S, et al. DIET@NET: Best Practice Guidelines for dietary assessment in health research. BMC Med 2017 Nov 15;15(1):202. [doi: 10.1186/s12916-017-0962-x] [Medline: 29137630]
44. 170.000 Lebensmittelprodukte auf dem deutschen Markt. Lebensmittelverband Deutschland. 2023. URL: https://www.lebensmittelverband.de/de/aktuell/20200731-tag-der-lebensmittelvielfalt [accessed 2025-01-22]

Abbreviations

DONALD study: Dortmund Nutritional and Anthropometric Longitudinally Designed study
DR: dietary record
EAN: European Article Number
GTIN: Global Trade Item Number
LEBTAB: LEBensmittelTABelle, food table
SAP SE: Systems, Applications & Products in Data Processing Societas Europaea
SUS: system usability scale
WDR: weighed dietary record

Edited by A Kushniruk; submitted 31.05.24; peer-reviewed by C Collins, D Budiningsari; revised version received 19.12.24; accepted 22.12.24; published 10.02.25.

Please cite as:
Klasen L, Koch SAJ, Benz ME, Conrad J, Alexy U, Blaszkiewicz K, Andone I, Nöthlings U
NutriDiary, a Smartphone-Based Dietary Record App: Description and Usability Evaluation
JMIR Hum Factors 2025;12:e62776
URL: https://humanfactors.jmir.org/2025/1/e62776
doi:10.2196/62776

Original Paper

Satisfaction and Usability of a Commercially Available Medication Adherence App (Medisafe) Among Medically Underserved Patients With Chronic Illnesses: Survey Study

Christa Hartch[1,2*], PhD, RN, FNP-C; Mary S Dietrich[2,3*], MS, PhD; B Jeanette Lancaster[4*], PhD, RN, FAAN; Shelagh A Mulvaney[2,5*], PhD; Deonni P Stolldorf[2*], PhD, RN

[1]School of Nursing and Health Sciences, Manhattanville University, Purchase, NY, United States

[2]School of Nursing, Vanderbilt University, Nashville, TN, United States

[3]Department of Biostatistics, Vanderbilt University Medical Center, Vanderbilt University, Nashville, TN, United States

[4]Sadie Heath Cabiness Professor and Dean Emerita, School of Nursing, University of Virginia, Charlottesville, VA, United States

[5]Department of Biomedical Informatics, Vanderbilt University Medical Center, Vanderbilt University, Nashville, TN, United States

[*]all authors contributed equally

Corresponding Author:
Christa Hartch, PhD, RN, FNP-C
School of Nursing and Health Sciences
Manhattanville University
2900 Purchase Street
Purchase, NY, 10577
United States
Phone: 1 914 323 5365
Email: christa.hartch@mville.edu

Abstract

Background: Research supports the use of mobile phone apps to promote medication adherence, but the use of and satisfaction with these apps among medically underserved patients with chronic illnesses remain unclear.

Objective: This study reports on the overall use of and satisfaction with a medication adherence app (Medisafe) in a medically underserved population.

Methods: Medically underserved adults who received care for one or more chronic illnesses at a federally qualified health center (FQHC) were randomized to an intervention group in a larger randomized controlled trial and used the app for 1 month (n=30), after which they completed a web-based survey. Objective data on app usage were provided as secondary data by the app company.

Results: The participants were very satisfied with the app, with all participants (30/30, 100%) somewhat or strongly agreeing that they would recommend the app to family and friends. Participants strongly agreed (28/30, 93%) that the reminders helped them remember to take their medications at the correct time each day, and they (28/30, 93%) found the app easy to use. Additional features accessed by some included educational features and the adherence report. Participants noted the helpfulness of having a medication list on their phones, and some used it during medication reconciliation at doctor visits. Use of the Medfriend feature, which alerts a social support person if a medication is missed, was low (n=2), but those who used it were very positive about the feature.

Conclusions: A commercially available medication adherence app was found to be useful by participants, and they were satisfied with the app and the additional features provided. The use of medication adherence mobile phone apps has the potential to positively influence chronic disease management in a medically underserved population on a large scale.

Trial Registration: ClinicalTrials.gov NCT05098743; https://clinicaltrials.gov/study/NCT05098743

(JMIR Hum Factors 2025;12:e63653) doi:10.2196/63653

KEYWORDS

medication adherence; mHealth; mobile phone; app; medically underserved; chronic disease; satisfaction; usage; health disparities

Introduction

Background

Medication adherence is vital for those with chronic illnesses who require long-term medication therapies to maintain optimal health. For example, medication adherence and persistence with high blood pressure medications are known to significantly decrease the risk of both cardiac events and stroke [1,2]. In those with type 2 diabetes, medication adherence with hypoglycemics reduces microvascular complications [3]. Unfortunately, the burden of chronic diseases is increasing, with an estimated 60% of adults in the United States having 1 chronic disease and 40% having 2 or more chronic diseases [4]. The growth of chronic disease burden coupled with a lack of medication adherence is associated with increased health care expenditures due to increased demands on health care resources [5,6] and poor health outcomes such as worsening disease status and even death [5]. The economic impact of low medication adherence is estimated to cost the US health care system between US $100 billion and $290 billion annually [5-8].

Medication adherence is therefore particularly important in medically underserved populations who seek care at federally qualified health centers (FQHCs). These centers serve communities and populations with a demonstrable unmet need for health services [9]. These centers are reporting growth in more complex patient populations because their patients have higher rates of chronic conditions and social risk factors associated with poorer health outcomes [10]. Additionally, lower rates of medication adherence are seen in lower socioeconomic populations [11,12] and those with multiple chronic conditions [13]. The reasons for this are influenced by social determinants of health and the material and social conditions in which people live [14]. Adverse social determinants of health are associated with lower medication adherence [15].

Mobile health (mHealth) interventions, defined as the use of mobile wireless technologies for public health [16], have been cited as a potential way to reduce health disparities among chronically ill and medically underserved populations [17,18]. However, despite the promise of these technologies, researchers indicate that mHealth interventions remain understudied in medically underserved populations [17,18]. This is true of medication adherence apps, which can support patients in adhering to their medications through reminders, medication educational information, adherence data, and social support. Studies have shown mixed results for the interest in mobile phone interventions in vulnerable populations [18,19]. Furthermore, research testing commercially available apps to manage chronic disease in a racially and ethnically diverse sample found that the usability of the tested apps in this population was suboptimal [20]. Understanding and gathering detailed data from diverse perspectives regarding the user experience of medication adherence apps will provide important information that is needed to support wider implementation.

A recent meta-analysis of the effectiveness of mobile apps on medication adherence in adults with chronic illnesses found that medication adherence mobile apps, which are designed to be used across a range of multiple chronic health conditions, remain underexplored [21]. This meta-analysis reported that in general, patients have a high acceptance of medication apps, but none of the studies analyzed included medically underserved populations [21]. Eight studies have demonstrated increased medication adherence with the use of medication adherence apps [22-29], but only 3 of these were conducted in low-income medically underserved populations [27-29]. Two of those studies in underserved populations were focused on hypertension [27,28], and the other included hypertension and type 2 diabetes and was a post hoc analysis of a digital health offering using a cluster-randomized design [29]. Only 1 of these studies, conducted in an urban low-income population with hypertension, obtained satisfaction information on the intervention [28]. Satisfaction with the app was high, and most participants felt they would use the app or a similar program in the future. Participants agreed that the app made it easier to keep track of their medications and that having a medication list on their phone made it easier to take care of themselves. More detailed feedback from the participants or information on which features of the app were used was not gathered [28].

A high-quality, free, commercially available smartphone medication adherence app called Medisafe supports patients in adhering to their medication regimen across disease states [30]. It uses a variety of advanced features, such as daily reminders, which can be snoozed, rescheduled, and marked as taken or missed; medication educational information in the form of medication cards and videos; an interaction checker; customizable refill reminders; adherence reports; and the ability to designate a social support person to be notified if a medication is skipped [23]. A randomized controlled trial (RCT) mixed methods evaluation in patients with coronary artery disease examined the efficacy [23] and the utility, acceptability, and engagement [31] of the Medisafe app. This study was conducted in a large urban tertiary hospital in Sydney, Australia and did not focus on a medically underserved population. In addition to improving self-reported medication adherence, overall utility was rated positively, with participants indicating that having their medication list on their phone and receiving timed reminders were useful. Most participants engaged with the app and its features; found the app acceptable, convenient, and easy to use; planned to continue using the app; and would recommend it to a family member or friend [31].

A qualitative study explored the potential benefits and barriers of using a mobile medication app in a medically underserved population in the United States [18]. The researchers found that patients were willing to try smartphone apps but expressed concerns about affordability, the technology being too complicated, not keeping phones with them all the time, and not being able to use all the features [18]. That study exposed a knowledge gap regarding the perceptions and user experiences of medically underserved patients with chronic illnesses who use free commercially available medication adherence apps.

Purpose

To address the knowledge gaps, a larger RCT investigating the efficacy of the Medisafe app (reported elsewhere) [32] was performed for evaluating the overall use and satisfaction of patients with a variety of chronic illnesses in a medically

underserved population in an FQHC in the United States. The efficacy portion of the RCT found significant improvements in both medication adherence (Cohen *d*=0.52; *P*=.01) and medication self-efficacy (Cohen *d*=0.43; *P*=.04) for participants assigned to use the app compared to the usual care group [32]. As part of this RCT, participants assigned to the intervention arm provided feedback and usage data regarding their experience using the Medisafe app [32]. This manuscript presents the summaries of the perceptions of patients enrolled in the intervention arm of the RCT regarding the usefulness of and satisfaction with the app features after 1 month of use. Given the improvements observed in medication adherence and self-efficacy, understanding patients perceived usefulness and satisfaction with the app is important to address potential barriers for uptake and use in larger medically underserved patient populations who often receive care for chronic illnesses in FQHCs.

Methods

Setting and Recruitment

Participants were recruited from November 2021 through June 2022 from an outpatient adult medicine department in an FQHC in the northeastern United States. The inclusion criteria for the RCT study were as follows: (1) adults aged 18 years or older, (2) having the ability to speak and understand English, (3) personally owning and using an Android smartphone (version 5.0 or above and at least 88 MB of phone space) or iOS smartphone (version 13 or later and at least 165 MB of phone space), and (4) taking at least one medication for a chronic condition based on the computerized medical record at the health center. Patients were excluded if they: (1) were already using a medication reminder app or other electronic reminder system such as phone alarms, (2) owned a smartphone not capable of downloading the app, (3) had a diagnosis of severe dementia or serious mental illness, or (4) were otherwise unable to use a mobile phone or the medication reminder software either physically or cognitively. For this study, only those participants who were randomized to the intervention group and used the Medisafe app were invited to participate in the survey.

Recruitment involved an informational flyer, a referral form from clinicians at the health center, and in-person recruitment. The flyer and referral forms were available to clinicians, staff, and patients in the FQHC offices and at the reception desk. The form contained study information, the contact information of the principal investigator (PI), and a place for patients interested in participation to provide their contact information. The form also contained a section for health care providers (HCPs) to refer potential patients and a section for their signature to verify that the patient's medications listed in the electronic health record were correct and current. The PI (CH) conducted in-person recruitment at the FQHC on multiple days per week and worked with clinic staff to identify potentially eligible patients. Although a convenience sample was used (ie, patients visiting the clinic on any given day), the risk of selection bias was reduced by using the aforementioned 2-prong approach to identify eligible patients for recruitment, inviting all patients meeting the eligibility criteria to participate, and using random assignment to either the intervention or control group. The 2 groups were not statistically different [32]. The PI approached eligible patients at the end of the health center visit to inform them of the study. Once the PI confirmed participant eligibility and obtained informed consent, participants were randomized to either the intervention or control group. Additional details of study procedures for the RCT have been previously published [32].

Statistical Analysis

Based on a preliminary efficacy study for the RCT [33], a total sample of 60 participants was estimated to enable the detection of differences between the groups with Cohen *d* effect values of 0.6-0.7 (80% power; α=.05) for the quantitative study variables [32]. As 30 participants were randomized to the app intervention group, their usage and satisfaction data are presented in this manuscript. Descriptive statistics and frequency distributions were used to describe the sample and determine if the data were normally distributed. Qualitative participant responses were transcribed into an Excel spreadsheet, and the content was coded and summarized as themes by the researcher (CH) and PhD faculty advisor (DPS).

Ethical Considerations

This research was approved by the Vanderbilt University Institutional Review Board (IRB #211409) and is registered with ClinicalTrials.gov (NCT05098743). All participants received a copy of the consent form. Based on participant preference, informed consent was completed as either an IRB-approved e-consent form or a hard copy. The consent form contained a privacy and confidentiality protection description ensuring that all study data are deidentified. Participants received a US $25 gift card after completion of the baseline survey and a US $35 gift card after completion of the follow-up survey.

Medisafe App Intervention

The Medisafe app is a Health Insurance Portability and Accountability Act (HIPAA) compliant medication adherence app that is available at no cost in the iTunes and Google app stores. In previous studies, Medisafe was ranked highly among medication reminder apps [30,34]. The Medisafe app provides interactive and customizable daily timed reminders to reinforce medication taking at a set time every day through a push notification, equivalent to an alarm or text message. The reminders can be snoozed, rescheduled, or marked as taken or skipped, and they are repeated a total of 3 times in 10-minute intervals if the participant does not mark the medication dose. Additional features include educational information in the form of a medication database that includes written and video content [30]. The written content is in the form of a medication card, which Medisafe terms a leaflet, and it reviews what the medicine is used for, medication interactions, what to do if the user misses a dose, what the user should watch for, possible side effects, how it should be used, and where to store the medication. Some medications also have video content, consisting of a brief clip of an HCP reviewing the most important considerations when taking the medication, which can be viewed on tapping the information icon. There is also an interactions tab that lists possible interactions with the medications or food/alcohol.

Lastly, there is an interaction checker where participants can check for interactions between their medications. The app also has a Medfriend feature, which allows participants to designate a family member or friend as their support person. The Medfriend feature will alert the designated Medfriend who can provide peer support and additional reminders through text messages, emails, or a telephone call if the patient misses a dose. The language mode of the app can be switched, if desired, to multiple foreign languages, including Spanish.

The PI helped consented participants set up the app using a copy of the patient's medication list extracted from the electronic health record. The PI also reviewed how participants could access and edit their medications; access medication educational content; and indicate when a medication was taken, skipped, or rescheduled. The PI reviewed with participants additional app features such as Medfriend, medication interaction checking, adherence reports, and refill reminders. Participants were also shown how to access the help and support section in the app. Following the app set up, the PI provided previously developed educational materials as a take-home resource. These materials, specific to either an iPhone or Android smartphone, included a laminated "quick tips" card with short instructions on the reviewed features and how to access them. Additional detailed instructions on how to use the app were printed in a question-and-answer format and distributed to participants.

Data Collection and Study Procedures

All study data were collected using observation, a REDCap web-based survey, and secondary data provided by the app company and were obtained using a data-sharing agreement between institutions. REDCap is a secure web-based software platform designed to support data capture for research studies [35,36]. Following consent, all participants completed the baseline study survey. Those randomized to the intervention group also completed a survey at study end to obtain feedback on the app, including usability and satisfaction.

Observation

While setting up the app for the intervention group, the PI completed a study-specific observational behavioral checklist. The purpose of the checklist was to inform the researchers if participants had difficulty setting up the app and how long it took them to do so. The checklist included documenting whether the participant had difficulty visualizing the app and had difficulty with dexterity while setting up the app, and mentioning the number of times the participant's input of medications needed to be corrected. The length of time in minutes from starting the download of the app to completing app set up and reviewing the app was also documented.

Survey

After 1 month, based on preference, participants completed the follow-up survey online, by phone, or in-person at the health center. Participants who did not complete the follow-up survey within 10 days of the 1-month follow-up date received 2 reminders via phone, email, or text message.

Measures

The end-of-study survey included 11 questions that assessed satisfaction and usability using a 5-point Likert scale with responses ranging from "strongly disagree" to "strongly agree." The survey questions were developed and pilot tested before use [33]. Seven of the first 11 questions were developed by Santo et al [31] and were used with permission in this study, while the remaining 4 were developed by the researchers. Six additional questions asked about the use of additional features such as the educational information, Medfriend feature, interaction checker, adherence report, refill reminder, and additional morning reminder of the Medisafe app. These questions asked participants whether they used a given feature, and if they did, whether they found the feature useful. There were open-text response options available to elicit qualitative data from the app participants such as how a feature helped them manage their medications and what they found most useful about the feature. The remainder of the survey included 6 general use questions previously developed and pilot tested, 4 of which were "Yes/No" questions (eg, did you use the refill reminder and did you have technical issues with the smartphone app?). The remaining 2 questions assessed how often medication reminders were received each day and which language the participant used.

Secondary Data

Deidentified usage data were obtained from the Medisafe company at study completion. Medisafe provided the PI with objective user interactions with the Medisafe app, such as whether educational information in the form of a leaflet was accessed by participants and whether the Medfriend feature was used.

Analysis

Descriptive statistics and frequency distributions were used to describe the sample and determine if the data were normally distributed. Open-ended survey question responses were imported into an Excel (Microsoft Corp) spreadsheet, and the content was coded and summarized as themes by the researcher (CH) and the PhD faculty advisor (DPS). This approach was taken given the short, free-text, and limited responses received. Since the qualitative data came from the open-ended survey responses, data collection was based on sample size rather than data saturation.

Results

Participant Characteristics

Complete details are included in Table 1, and information can also be found in the app efficacy manuscript [32]. A flowchart of study participants in the main RCT can be found in Figure 1. The median age of the 30 participants using the app was 53.5 years (IQR 37-76 years). Most participants in the intervention group were non-White (23/29, 79%). Races/ethnicities were as follows: Asian (5/29, 17%), Black or African American (10/29, 35%), Hispanic/Latino (4/29, 14%), Native American or Alaska Native (1/29, 3%), and other (3/29, 10%). More than 75% of the participants had government insurance (25/30, 83%), and a small number of participants were uninsured (2/30, 7%). Brief

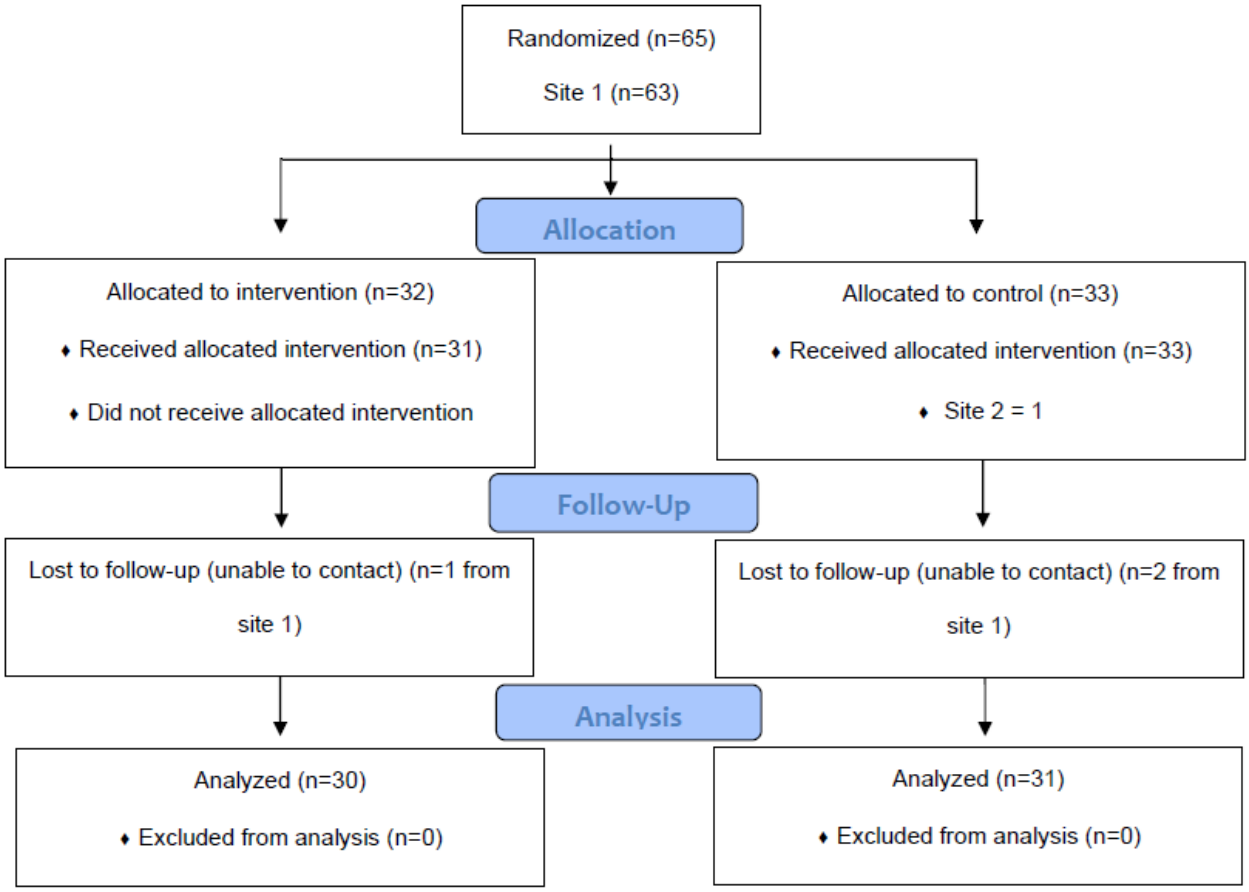
health literacy scores were generally high (median 12.0 of a possible score of 15, IQR 5-15). Slightly more than half of the participants (16/29, 55%) reported that it was either very or somewhat difficult to pay their monthly bills. The most common chronic illness was hypertension (22/30, 73%), followed by hyperlipidemia (19/30, 63%) and type 2 diabetes (14/30, 47%). Most participants (25/30, 83.3%) had 2 or more chronic illnesses.

Table 1. Sociodemographic characteristics of the participants using the app.

Characteristic	Value, n (%)
Race/ethnicity (N=29)	
Asian	5 (17)
Black or African American	10 (35)
Hispanic/Latino	4 (14)
Native American or Alaska Native	1 (3)
White	6 (21)
Other	3 (10)
Marital status (N=30)	
Married/partnered	15 (50)
Single/never married	15 (50)
Employment status (N=30)	
Employed	15 (50)
Unemployed	12 (40)
Retired	3 (10)
Education (N=30)	
Some high school or less	8 (27)
High school graduate	4 (13)
College credit, no degree	8 (27)
Trade/vocational training	4 (13)
Associate’s degree	2 (7)
Bachelor’s degree or higher	4 (13)
Difficulty paying bills (N=29)	
Very difficult	6 (21)
Somewhat difficult	10 (35)
Not very difficult	8 (28)
Not at all difficult	5 (17)
Type of health insurance (N=30)	
Uninsured (sliding scale)	2 (7)
Government insurance	25 (83)
Private insurance	3 (10)
Current chronic illness (N=30)	
Hypertension	22 (73)
Type 2 diabetes	14 (47)
Hyperlipidemia	19 (63)
Asthma	5 (17)
Other[a]	11 (37)

[a]Includes depression (n=1), type 1 diabetes (n=1), chronic obstructive pulmonary disease (n=1), heart disease (n=1), cirrhosis (n=1), anxiety (n=1), gout (n=1), rheumatoid arthritis (n=1), thyroid disorder (n=1), hypothyroidism (n=2), gastroesophageal reflux disease (n=1), arthritis (chronic pain) (n=1), and fibromyalgia (n=1).

Figure 1. Flowchart of study participants in the larger randomized controlled trial of the efficacy of the app intervention.



Behavioral Observations While Setting Up the App

During app set up, some participants (4/30, 13%) expressed difficulty visualizing the app owing to the unavailability of their eyeglasses, which they stated were either in their car or left at home. No participants had difficulty with dexterity while setting up the app, and the median time it took from starting the app download to completing the set up and reviewing the app was 15 minutes (IQR 10.0-25.0; minimum 10, maximum 30 minutes). The majority of participants (21/30, 70%) did not need to be corrected when they entered the medications. However, 4 (13%) were corrected by the researcher once, 3 (10%) were corrected twice, and 2 (7%) were corrected thrice. Patients were corrected when they spelled the medication name incorrectly, chose the incorrect medication dose, or set an incorrect time for the reminder.

Satisfaction and Utility

Summaries of the participants' reports of satisfaction are presented in Table 2.

Most participants (27/30, 90%) strongly agreed that they liked the app design, while most (25/30, 83%) strongly agreed that it was useful to have their medication list on their smartphone. Some participants (2/30, 7%) mentioned the usefulness of the app when seeing other HCPs to indicate the medications they were taking during medication reconciliation. Furthermore, a large proportion of participants (28/30, 93%) strongly agreed that the reminders helped them to remember to take their medications at the correct time each day. The majority of participants strongly agreed that the app was easy to use (27/30, 90%) and convenient (28/30, 93%) and that they would continue using the app (26/30, 87%). A small number of participants (2/30, 7%) somewhat agreed that they would continue using the app. It is important to note that some participants (2/30, 7%) strongly disagreed that they would continue using the app, because they found the reminders annoying. All the participants (30/30, 100%) somewhat agreed or strongly agreed that they would recommend the app to family and friends.

Table 2. Satisfaction with the app (N=30).

Satisfaction information	Value (N=30), n (%)
Liked the app design	
Neutral	1 (3)
Somewhat agree	2 (7)
Strongly agree	27 (90)
It is easy to tap the correct icon with my finger	
Neutral	1 (3)
Somewhat agree	1 (3)
Strongly agree	28 (93)
I am able to see all the options in the app	
Neutral	3 (10)
Somewhat agree	1 (3)
Strongly agree	26 (87)
It is useful to have a medication list on the smartphone	
Somewhat disagree	1 (3)
Neutral	1 (3)
Somewhat agree	3 (10)
Strongly agree	25 (83)
Reminders helped me remember to take my medications at the correct time each day	
Neutral	1 (3)
Somewhat agree	1 (3)
Strongly agree	28 (93)
Found it easy to use the app	
Somewhat disagree	1 (3)
Neutral	1 (3)
Somewhat agree	1 (3)
Strongly agree	27 (90)
Found it easy to set up reminders in the app	
Neutral	3 (10)
Somewhat agree	2 (7)
Strongly agree	25 (83)
Found it convenient to have the app	
Strongly disagree	1 (3)
Somewhat disagree	1 (3)
Strongly agree	28 (93)
Found it useful to snooze the reminder	
Strongly disagree	1 (3)
Somewhat disagree	1 (3)
Neutral	10 (33)
Somewhat agree	3 (10)
Strongly agree	15 (50)
Will continue using the app	
Strongly disagree	2 (7)

Satisfaction information	Value (N=30), n (%)
Somewhat disagree	1 (3)
Neutral	1 (3)
Strongly agree	26 (87)
Will recommend the app to family and friends	
Somewhat agree	2 (7)
Strongly agree	28 (93)

The 2 participants (10%) who were dissatisfied with the app described the reasons why. One said:

> *It is annoying when you get a reminder and you are in the middle of doing your work. This is not for everybody. I work on the computer and on my cell phone and it is very distracting to receive the reminder in the middle of working on something. It might be better for someone who doesn't have as much going on. I find it very distracting.*

The other participant had technical difficulties but blamed it on the type of phone they had:

> *The bug thing with the notification alarms was a problem. I have an android - a cheap phone. My phones get destroyed because of the type of work I do.*

Use of Educational Information

Almost half of the participants (12/30, 40%), self-reported accessing educational information. To the contrary, objective usage data from Medisafe indicated that only 3 participants (10%) accessed the educational content, which was defined as cards termed “leaflets” or videos. According to Medisafe, this was done for a total of 14 medications. Additionally, Medisafe reported that only 1 participant accessed 4 different videos and 1 leaflet, 1 participant accessed 1 leaflet, and 1 participant accessed 6 different videos and 2 leaflets. Although not all participants actually accessed the information, those who reported accessing the educational information (12/12, 100%) found the information useful. When asked about how they used the educational information in the app, the participants reported learning about the side effects of the medications (6/12, 50%), reported that it was helpful for general knowledge (4/12, 33%), and mentioned using it to learn more about medication and food interactions (2/12, 17%).

Medfriend Feature

Based on usage data from Medisafe, only 1 participant (3%) used the Medfriend feature. That participant was very positive about the feature and reported that her husband would call her to say, “Are you taking your medicines?” She stated:

> *It gets him involved. It makes him recognize that I need support and I need to take the medicine. It makes me know he loves me.*

Another participant self-reported using the Medfriend feature, but there was no indication of use in the Medisafe data. The participant reported that when her husband was notified, he would send a text about her forgetting her medications and she would remember to take them. Those who did not use the Medfriend option were asked, “who might that person be for you?” Among those who responded (18/30, 60%), the top 3 most common responses were their sibling (4/18, 22%), their child (4/18, 22%), and their husband or wife (4/18, 22%).

Interaction Checker

All participants who used the interaction checker (5/30, 17%) agreed that it was useful. One participant reported that 2 of the medications she had been taking together should be taken separately and stated, “It was a lifesaver!” The other 4 participants expressed an appreciation for being able to have access to this type of information. The use of the interaction checker was distinct from the educational content and could not be verified in the Medisafe data as Medisafe does not register or track the use of the interaction checker.

Adherence Reports and Reminders

Participants (8/30, 27%) who checked their adherence report agreed it was useful. The adherence report provided them with a history of their daily missed and taken medications as well as a weekly adherence percentage based on what they reported when marking medication reminders in the app. Participants reported experiencing positive reinforcement for adhering to their medications through the adherence report, mentioned the affirmation they received when they saw a high percentage of adherence, and reported appreciating the positive reinforcement as useful. Some participants (2/30, 7%) mentioned that it incentivized them to reach higher levels of adherence.

Slightly less than half of the participants (13/30, 43%) received reminders to take their medications 2 times a day, while around one-third (9/30, 30%) received reminders 3 or more times a day. Reminders were generally well received:

> *I like the reminder. The shaking of the pill bottle helps me. Sometimes I will wake up at night and remember hearing the shaking pill bottle that day and I will get out of bed and check if I took my pills that day. I might be cooking with the grandkids and the first reminder goes off. I might ignore it but with the second reminder I might put the bottles on the counter so I can remember.*

Three participants (15%) mentioned the app’s helpfulness, particularly for those who have multiple chronic illnesses and take multiple medications.

> *This app was a lifesaver. I take a lot of different medication so sometimes I forget whether I took the*

> *medication or not. I can check the app to see if I took it or not.*

One of those 3 participants commented as follows:

> *It is perfect for people who have multiple illnesses and take a lot of different medications. I take eight different medications a day.*

Most of the participants (27/30, 90%) did not use the refill reminder. In their comments, a number of participants said they received automatic refill reminders from their pharmacy and therefore did not need this feature of the app. All participants (30/30, 100%) used the app in English, and the majority (24/30, 80%) did not make any changes, such as changing the time of a reminder, removing or adding a medication, or changing the medication dose in the app. Many participants (13/30, 43%) said they would use the app to manage someone else's medications.

Technical Difficulties Using the App

Some participants (4/30, 13%) mentioned they had technical difficulties. Of those who reported difficulties, 3 (75%) needed to allow notifications from the app to hear the reminders. Another user mentioned that they had to tap the "take all" icon a number of times before it registered and suggested that it should be made bigger or be more centrally placed. Participants gave additional feedback about the app when asked (20/30, 67%). In this section, participants (7/20, 35%) specifically mentioned liking the reminder.

> *It was really nice to hear that shaking sound. It was fun.*

Some participants (2/20, 10%) reported that the snooze function was particularly helpful when they were away from home.

> *The snooze option is helpful to use when I am out and don't have my medications. When I come back home it reminds me so I remember to take it.*

Social Support

The 2 participants (7%) who self-reported using the Medfriend feature were very enthusiastic.

> *It's a great app and I love it. My husband is on it for his meds and I am his Medfriend. I am also going to get my mother hooked up on it.*

The other participant shared her thoughts about the feature, highlighting her increased feelings of self-efficacy and social support.

> *This app is about being a team player. You are able to help me and I am able to help you. I can now say "I did it" "I can do this." This is a good app. I can't see anyone who is interested in their health not using this app. Since being introduced to this app I know that it is there for me.*

Discussion

Principal Findings

As part of an RCT using the Medisafe medication adherence app in a medically underserved population with a variety of chronic illnesses, behavioral observations on app use and satisfaction and usage data were gathered from participants in the intervention arm of the study and the Medisafe company. The quantitative RCT results (reported elsewhere) [32] found significant improvements in both medication adherence and medication self-efficacy for participants who used the app. The portion of the RCT presented in this manuscript, which collected behavioral observations and satisfaction and usage data from the intervention arm, identified that participants were satisfied with the app and found it useful. Even though the use of the additional features was generally low, those who used them found them useful. Most participants did not need help setting up the app. An important strength of this RCT is that it explored patients' perceptions of the usefulness of the app and their satisfaction with the app and therefore fills an important knowledge gap. This was done by collecting both quantitative and qualitative data through open-response questions, which gave voice to the perspectives of a low-income racially and ethnically diverse sample of adults with chronic illnesses receiving care in an FQHC [17,18]. As FQHCs are reporting a growth in the rates of treating complex chronic conditions [10] and there are lower rates of medication adherence among populations with lower socioeconomic status and those with multiple conditions [37], the implementation of tools to enhance medication adherence is imperative. Understanding the user experience with the Medisafe mobile app demonstrated that wider-scale use of the Medisafe app is feasible in a low-income population with multiple chronic illnesses. Systematic literature reviews have pointed out a gap in implementation studies of mobile app interventions in this population [38]. This study addressed an important knowledge gap by demonstrating that the use of a commercially available free medication adherence app is a viable option for medically underserved adult patients with chronic diseases.

Prior research found that medically underserved patients expressed reluctance about paying for a medication adherence app [18]. While not directly addressed in this study, some participants anecdotally asked before enrolling in the study if they would need to pay for the app, and when told it was free, they expressed interest in participating. This underscores the importance of not having patients incur additional costs for the technology and was a strength of this intervention.

Most participants were able to set up the app with minimal assistance, with a median duration time of 15 minutes during the behavioral observation. It is important to note that 30% of the participants needed to be corrected 1-3 times when setting up the app, thus pointing to the importance of helping some patients set up the app initially and checking that the medications are entered accurately. This highlights a difficulty with individuals setting up the app. Although not implemented in this study, another option is to import medications from other databases, such as Apple Health, or a pharmacy directly. This may shorten the time it takes to set up the app. Once the app is set up properly, in addition to HCPs assisting patients, the help and support page and the company contact could serve as a resource for patients.

Survey data indicated that satisfaction with the app was high, with most patients strongly agreeing it was easy to use. All

intervention group participants (30/30, 100%) strongly agreed that they would recommend the app to family or friends. This was higher than the proportion in the study by Santo et al [31], which used Medisafe and found that 78.6% of patients with coronary heart disease would recommend it [31], and the study by Anglada-Martinez et al [24], which used a similar app and found that 71.4% of patients receiving treatment for hypertension, dyslipidemia, heart failure, or HIV would recommend it [24]. Both these studies were, however, conducted outside of the United States. It may be that Americans are more familiar with app technologies and feel more comfortable recommending apps to others.

Qualitative research conducted in a medically underserved population with chronic illness regarding the use of medication adherence apps found that technical issues and complexity were concerns when setting up and using these apps [18]. One study involving a medication adherence app similar to Medisafe indicated that 50% of participants reported problems receiving reminders [24]. The Medisafe app used in this study is a commercially available app with high-quality assessment ratings [30,39], and in this study, technical issues were rare. The most common issue was not receiving the reminders until the participant allowed notifications from the app in their phone settings. No participants expressed that the app was technically complex, which was previously cited as a concern in this population but was not an issue in this study [18]. One participant had ongoing technical issues receiving reminders. These findings suggest that the Medisafe app can be implemented in this population from a technology standpoint, and participants did not find it difficult to use.

Similar to other studies involving the Medisafe app, feedback results point to receiving timed medication reminders as the most used aspect of the app [31]. Furthermore, feedback regarding the app aligns with the findings of other studies linking medication reminders with medication adherence [22,25,28,29]. Participants found the snooze function of the reminder helpful when they were not home to take their medications and used this function as a reminder to take their medications when they got home. The snooze function therefore was an important component of the app when participants experienced disruptions in their routines, such as being away from their medications. The findings also align with a previous study in a medically underserved population where participants indicated that disruptions in their daily routines negatively affected their medication adherence [18]. This study demonstrated that patients used the reminder feature when available, and the majority of patients found it helpful in improving medication adherence. The 1 participant who was bothered by the reminders used his phone for work and found the reminders distracting if he was using his phone for work purposes. The reminders predominantly targeted the phenomenon of forgetting, which has been found by a study to be the most likely cause of reported nonadherence in low-income uninsured patients with multiple chronic illnesses [13].

Research has shown that both patient knowledge of medications and their satisfaction with the information provided about their medications can improve medication adherence [40]. There was a discrepancy between the data reported by Medisafe and the number of participants who self-reported accessing educational information. Although the reason for this discrepancy is not clear, several possibilities exist. First, participants could have overstated the use of educational features. Another potential reason might be related to the specific data Medisafe defines as educational data. Medisafe does not collect data on the use of the interaction tracker or the “For You” tab at the bottom of the app and only collects data if a participant clicks on the educational leaflet and opens it up. In contrast, participants might have perceived content under the “For You” tab and drug interaction materials or videos as educational materials, resulting in a discrepancy between patient self-reported data and Medisafe data regarding accessing educational content. The feedback received demonstrated that participants who reported accessing the educational information (less than half) were very positive about doing so. The educational information was found to be useful for learning about side effects and food and medication interactions. Because individuals have different preferences for the amount of medication information they receive and the way that information is delivered [41], the modularity of the Medisafe app is useful to facilitate patient education in a practical and less burdensome way. The information is available at the patient’s fingertips and can be accessed as frequently as needed to learn what they want at their convenience. The educational app feature is also advantageous to HCPs as it alleviates some of the burden and time commitment associated with educating patients about their medications.

Social support has been found to have a positive effect on medication adherence [42-44]. Studies deploying digital technologies in the form of web-based online communities to provide social support have generally demonstrated that they can support people emotionally, socially, practically, and politically [45]. However, using technology to provide social support has not been studied in the context of a commercially available medication adherence mobile app. This study addressed this gap by studying social support via a commercially available app in the context of medication adherence. The Medisafe app offers social support in the form of Medfriend, and this is the first known study to incorporate this feature as part of the study intervention. Some studies have pointed out that online social support networks for those with specific chronic illnesses lessened the burden on relationships with family and friends, who are referred to as “offline” support persons [45,46]. However, despite asking and offering to demonstrate how to set up the Medfriend feature, usage of the Medfriend feature was very low. This study did not gather information on why participants chose not to set up the Medfriend feature, and this is a limitation of the study. It may have been because this feature was seen as too burdensome by the patients or their support people, most of whom were identified as family members. Patients might have avoided using Medfriend due to confidentiality concerns associated with this feature, which entails giving access to the user’s medication list, as many participants (13/30, 43%) were willing to use the app to manage someone else’s medications but chose not to share their own medication information. Another challenge regarding the Medfriend feature was that there was a discrepancy between Medisafe data and self-reported data on the use of the Medfriend feature. Despite the aforementioned concerns, the 2

participants who reported using the Medfriend feature were satisfied with it as they perceived that the app fostered social support. To further explore the social support feature of the app, research on dyads who use the app to manage the medications of family members might shed new light on the phenomenon of incorporating social support into mobile apps. By studying a subset of the population, including patients and their caregivers, the social support feature may be used more frequently. If the confidentiality of medication lists proves to be a barrier, a feature that dissociates specific medications from the reminder might address that concern. Support persons could receive a general text that their online Medfriend has not taken their medications without sharing details on the specific medications.

Participants reported that having a list of medications on their phone was beneficial, which was also noted in a population of patients with coronary heart disease who used the Medisafe app in Australia [31]. When managing chronic illnesses, patients are often referred outside of the FQHC setting or require hospitalization to receive care. Some participants mentioned using the phone medication list for medication reconciliation when seeing other HCPs. This finding is in contrast to that of another study of patients presenting to an emergency department setting, which found that emergency department patients rarely used their mobile phones to share their medication list during medication reconciliation [47]. Medication reconciliation can be facilitated through the adoption of these technologies. HCPs in both primary care and tertiary care settings should suggest and support patients with implementing researched medication adherence mobile apps. The sample of this study included many patients with multiple chronic conditions. These participants appreciated the ability of the app to work across multiple chronic illnesses and its helpfulness when taking multiple medications. This finding underscores the importance of advocating for the use of medication adherence apps like Medisafe, which can work across a range of illnesses and medications and can be easily adjusted when medications change over the disease trajectory. Additionally, participants who used the adherence report felt that it provided positive reinforcement and was an incentive to reach higher levels of adherence. Similar to what has been reported in other studies, a majority of participants reported not using the refill reminder because they already received text alerts from their pharmacy, which they found helpful [18]. When HCPs select apps for patients to enhance their medication adherence, careful attention to app features and evaluation of existing research findings, such as the findings of this study, are important.

Limitations

This study had several limitations. First, the study duration of 1 month does not provide insights regarding long-term patient satisfaction and continued use of app features, which are important aspects of chronic disease management. Though this study found high satisfaction and usability of the app during the first month of use, future studies should evaluate the role of time in app usage and satisfaction. Medisafe data and self-reported data showed that the uptake of educational information and the Medfriend social support feature was low. There was an unresolved discrepancy in the number of participants who reported accessing educational information and the actual usage identified from the Medisafe data. The discrepancy might be because Medisafe data only captured if the leaflet was accessed. Patients may have perceived accessing educational information as clicking the interaction button or clicking the “For You” tab at the bottom of the app, which Medisafe data did not capture. This can be clarified in the future by a more detailed definition of what constitutes the educational features of the app. Incorporating interviews to clarify subsequent survey results would strengthen future research studies. Another limitation of the study is that we did not gather participants’ inputs about why they chose not to use the Medfriend feature. Therefore, this study cannot speak about the potential benefits of this feature. Finally, although the app can be used in several languages and many patients who seek care at FQHCs speak a primary language other than English [48], the researchers were not able to incorporate multiple languages into the study protocol.

Future Research

FQHCs and primary care settings working with adults who are chronically ill should consider medication adherence mobile phone apps as acceptable and practical tools to support medication adherence. Future studies could include a larger sample, consider the use of the available provider portal, and consider the experiences of both providers and patients. Cost analysis could be performed, and hospitalization rates and long-term usage and health outcomes over time could be studied. This study was for a 30-day period, but a study with a longer duration is necessary to see if the use of the app is sustained over time. In this study, only 2 participants reported barriers to using the app, and a larger long-term study could further explore barriers to sustained use and strategies for maintaining engagement in this population. Future research should use mixed methods to provide insights into app modification, the nature of barriers to use, and how app features, such as the Medfriend feature, could more easily be implemented among patients who might benefit the most from such features. As uptake of the additional features of the app, such as educational information and the Medfriend option, was low in this study, future research using larger datasets could explore what types of patients chose to use specific features and why they did. We purposefully did not require certain features to be used because we wanted to organically discover which features were most often used, if any.

Studying the usability of the app and its associated effects in ethnic populations in various languages is an important area of future research as community health centers serve a large number of patients with limited English proficiency [48].

Conclusion

This study demonstrated that the medication adherence app is a useful, convenient, and feasible intervention in an FQHC setting. The various features of this app positively influenced the medication-taking behaviors of adults with one or more chronic illnesses. Participants were satisfied with the app and the features they chose to use. Reminders were viewed as helpful by the majority of participants. The medication list feature was particularly useful for patients who had multiple chronic

conditions and saw multiple providers, and some used it to facilitate medication reconciliation. The findings of this study have important clinical implications, as clinicians can recommend the use of medication adherence apps as tools to provide support in adhering to medication regimens and as additional tools to use during medication reconciliation.

Acknowledgments

We would like to extend our gratitude to the leadership and staff of the health center and the participants who made this work possible. Support for this study was provided by the Vanderbilt Institute for Clinical and Translational Research (VICTR) #VR56076, Student Achievement Research Award (SARA), and VICTR #VR60349. The research study was also supported by CTSA award UL1TR002243 from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences of the National Institutes of Health.

Data Availability

Limited deidentified data are available upon request from the corresponding author.

Conflicts of Interest

None declared.

References

1. Corrao G, Parodi A, Nicotra F, Zambon A, Merlino L, Cesana G, et al. Better compliance to antihypertensive medications reduces cardiovascular risk. J Hypertens 2011 Mar;29(3):610-618. [doi: 10.1097/HJH.0b013e328342ca97] [Medline: 21157368]
2. Vrijens B, Antoniou S, Burnier M, de la Sierra A, Volpe M. Current situation of medication adherence in hypertension. Front Pharmacol 2017 Mar 01;8:100 [FREE Full text] [doi: 10.3389/fphar.2017.00100] [Medline: 28298894]
3. Yu AP, Yu YF, Nichol MB. Estimating the effect of medication adherence on health outcomes among patients with type 2 diabetes--an application of marginal structural models. Value Health 2010 Dec;13(8):1038-1045 [FREE Full text] [doi: 10.1111/j.1524-4733.2010.00787.x] [Medline: 20946182]
4. Chronic Diseases in America. National Center for Chronic Disease Prevention and Health Promotion. URL: https://www.cdc.gov/chronic-disease/about/index.html [accessed 2024-12-26]
5. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/nejmra050100]
6. Viswanathan M, Golin CE, Jones CD, Ashok M, Blalock SJ, Wines RC, et al. Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review. Ann Intern Med 2012 Dec 04;157(11):785-795 [FREE Full text] [doi: 10.7326/0003-4819-157-11-201212040-00538] [Medline: 22964778]
7. Thinking outside the pillbox: A system-wide approach to improving patient medication adherence for chronic disease. New England Healthcare Institute. 2009. URL: https://www.akleg.gov/basis/get_documents.asp?session=28&docid=3370 [accessed 2024-12-26]
8. Peterson A, Takiya L, Finley R. Meta-analysis of trials of interventions to improve medication adherence. Am J Health Syst Pharm 2003 Apr 01;60(7):657-665. [doi: 10.1093/ajhp/60.7.657] [Medline: 12701547]
9. Medically Underserved Areas and Population (MUA/P). Health Resources and Services Administration. URL: https://bhw.hrsa.gov/workforce-shortage-areas/shortage-designation#mups [accessed 2024-12-26]
10. Health Centers are Providing Care to Growing Numbers of Patients with Complex Needs. National Association of Community Health Centers. URL: https://opus-nc-public.digitellcdn.com/uploads/nachc/articles/808aaa175320895896c418d5fa27d6f6.pdf [accessed 2024-12-26]
11. Osborn CY, Mayberry LS, Wagner JA, Welch GW. Stressors may compromise medication adherence among adults with diabetes and low socioeconomic status. West J Nurs Res 2014 Oct 25;36(9):1091-1110 [FREE Full text] [doi: 10.1177/0193945914524639] [Medline: 24569697]
12. Bidwal M, Lor K, Yu J, Ip E. Evaluation of asthma medication adherence rates and strategies to improve adherence in the underserved population at a Federally Qualified Health Center. Res Social Adm Pharm 2017 Jul;13(4):759-766. [doi: 10.1016/j.sapharm.2016.07.007] [Medline: 27595427]
13. Fernandez-Lazaro CI, Adams DP, Fernandez-Lazaro D, Garcia-González J, Caballero-Garcia A, Miron-Canelo JA. Medication adherence and barriers among low-income, uninsured patients with multiple chronic conditions. Res Social Adm Pharm 2019 Jun;15(6):744-753. [doi: 10.1016/j.sapharm.2018.09.006] [Medline: 30241872]
14. Reutter L, Kushner KE. 'Health equity through action on the social determinants of health': taking up the challenge in nursing. Nurs Inq 2010 Sep 16;17(3):269-280. [doi: 10.1111/j.1440-1800.2010.00500.x] [Medline: 20712665]

15. Wilder ME, Kulie P, Jensen C, Levett P, Blanchard J, Dominguez LW, et al. The impact of social determinants of health on medication adherence: a systematic review and meta-analysis. J Gen Intern Med 2021 May 29;36(5):1359-1370. [doi: 10.1007/s11606-020-06447-0] [Medline: 33515188]
16. mHealth: use of appropriate digital technologies for public health: report by the Director-General. World Health Organization. 2017. URL: https://iris.who.int/handle/10665/274134 [accessed 2024-12-26]
17. Peek ME. Can mHealth interventions reduce health disparities among vulnerable populations? Divers Equal Health Care 2017;14(2):44-45. [doi: 10.21767/2049-5471.100091]
18. Hincapie AL, Gupta V, Brown SA, Metzger AH. Exploring perceived barriers to medication adherence and the use of mobile technology in underserved patients with chronic conditions. J Pharm Pract 2019 Apr 06;32(2):147-153. [doi: 10.1177/0897190017744953] [Medline: 29212410]
19. Stockman M, Modzelewski K, Steenkamp D. Mobile health and technology usage by patients in the diabetes, nutrition, and weight management clinic at an urban academic medical center. Diabetes Technol Ther 2019 Jul;21(7):400-405. [doi: 10.1089/dia.2018.0369] [Medline: 31045447]
20. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec 14;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
21. Peng Y, Wang H, Fang Q, Xie L, Shu L, Sun W, et al. Effectiveness of mobile applications on medication adherence in adults with chronic diseases: a systematic review and meta-analysis. J Manag Care Spec Pharm 2020 Apr;26(4):550-561 [FREE Full text] [doi: 10.18553/jmcp.2020.26.4.550] [Medline: 32223596]
22. Morawski K, Ghazinouri R, Krumme A, Lauffenburger JC, Lu Z, Durfee E, et al. Association of a smartphone application with medication adherence and blood pressure control: the MedISAFE-BP randomized clinical trial. JAMA Intern Med 2018 Jun 01;178(6):802-809 [FREE Full text] [doi: 10.1001/jamainternmed.2018.0447] [Medline: 29710289]
23. Santo K, Singleton A, Rogers K, Thiagalingam A, Chalmers J, Chow CK, et al. Medication reminder applications to improve adherence in coronary heart disease: a randomised clinical trial. Heart 2019 Feb 27;105(4):323-329. [doi: 10.1136/heartjnl-2018-313479] [Medline: 30150326]
24. Anglada-Martínez H, Martin-Conde M, Rovira-Illamola M, Sotoca-Momblona JM, Sequeira E, Aragunde V, et al. Feasibility and preliminary outcomes of a web and smartphone-based medication self-management platform for chronically ill patients. J Med Syst 2016 Apr 12;40(4):99. [doi: 10.1007/s10916-016-0456-y] [Medline: 26872781]
25. Labovitz DL, Shafner L, Reyes Gil M, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419. [doi: 10.1161/strokeaha.116.016281]
26. Svendsen M, Andersen F, Andersen K, Pottegård A, Johannessen H, Möller S, et al. A smartphone application supporting patients with psoriasis improves adherence to topical treatment: a randomized controlled trial. Br J Dermatol 2018 Nov 05;179(5):1062-1071. [doi: 10.1111/bjd.16667] [Medline: 29654699]
27. Chandler J, Sox L, Kellam K, Feder L, Nemeth L, Treiber F. Impact of a culturally tailored mHealth medication regimen self-management program upon blood pressure among hypertensive Hispanic adults. Int J Environ Res Public Health 2019 Apr 06;16(7):1226 [FREE Full text] [doi: 10.3390/ijerph16071226] [Medline: 30959858]
28. Patel S, Jacobus-Kantor L, Marshall L, Ritchie C, Kaplinski M, Khurana PS, et al. Mobilizing your medications: an automated medication reminder application for mobile phones and hypertension medication adherence in a high-risk urban population. J Diabetes Sci Technol 2013 May 01;7(3):630-639 [FREE Full text] [doi: 10.1177/193229681300700307] [Medline: 23759395]
29. Moorhead P, Zavala A, Kim Y, Virdi NS. Efficacy and safety of a medication dose reminder feature in a digital health offering with the use of sensor-enabled medicines. J Am Pharm Assoc (2003) 2017 Mar;57(2):155-161.e1. [doi: 10.1016/j.japh.2016.12.067] [Medline: 28159505]
30. Santo K, Richtering SS, Chalmers J, Thiagalingam A, Chow CK, Redfern J. Mobile phone apps to improve medication adherence: a systematic stepwise process to identify high-quality apps. JMIR Mhealth Uhealth 2016 Dec 02;4(4):e132 [FREE Full text] [doi: 10.2196/mhealth.6742] [Medline: 27913373]
31. Santo K, Singleton A, Chow CK, Redfern J. Evaluating reach, acceptability, utility, and engagement with an app-based intervention to improve medication adherence in patients with coronary heart disease in the MedApp-CHD study: a mixed-methods evaluation. Med Sci (Basel) 2019 Jun 04;7(6):68 [FREE Full text] [doi: 10.3390/medsci7060068] [Medline: 31167489]
32. Hartch CE, Dietrich MS, Lancaster BJ, Stolldorf DP, Mulvaney SA. Effects of a medication adherence app among medically underserved adults with chronic illness: a randomized controlled trial. J Behav Med 2024 Jun 21;47(3):389-404 [FREE Full text] [doi: 10.1007/s10865-023-00446-2] [Medline: 38127174]
33. Hartch CE, Dietrich MS, Stolldorf DP. Effect of a medication adherence mobile phone app on medically underserved patients with chronic illness: preliminary efficacy study. JMIR Form Res 2023 Dec 11;7:e50579 [FREE Full text] [doi: 10.2196/50579] [Medline: 38079192]
34. Morawski K, Ghazinouri R, Krumme A, McDonough J, Durfee E, Oley L, et al. Rationale and design of the medication adherence improvement support app for engagement-blood pressure (MedISAFE-BP) trial. Am Heart J 2017 Apr;186:40-47. [doi: 10.1016/j.ahj.2016.11.007] [Medline: 28454831]

35. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
36. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
37. Rolnick SJ, Pawloski PA, Hedblom BD, Asche SE, Bruzek RJ. Patient characteristics associated with medication adherence. Clin Med Res 2013 Jun 11;11(2):54-65 [FREE Full text] [doi: 10.3121/cmr.2013.1113] [Medline: 23580788]
38. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the united states: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
39. Park JYE, Li J, Howren A, Tsao NW, De Vera M. Mobile phone apps targeting medication adherence: quality assessment and content analysis of user reviews. JMIR Mhealth Uhealth 2019 Jan 31;7(1):e11919 [FREE Full text] [doi: 10.2196/11919] [Medline: 30702435]
40. Alhewiti A. Adherence to long-term therapies and beliefs about medications. Int J Family Med 2014 Feb 13;2014:479596-479598 [FREE Full text] [doi: 10.1155/2014/479596] [Medline: 24688792]
41. Horne R, Hankins M, Jenkins R. The Satisfaction with Information about Medicines Scale (SIMS): a new measurement tool for audit and research. Qual Health Care 2001 Sep 01;10(3):135-140 [FREE Full text] [doi: 10.1136/qhc.0100135] [Medline: 11533420]
42. DiMatteo MR. Social support and patient adherence to medical treatment: a meta-analysis. Health Psychol 2004 Mar;23(2):207-218. [doi: 10.1037/0278-6133.23.2.207] [Medline: 15008666]
43. Molloy GJ, Perkins-Porras L, Bhattacharyya MR, Strike PC, Steptoe A. Practical support predicts medication adherence and attendance at cardiac rehabilitation following acute coronary syndrome. J Psychosom Res 2008 Dec;65(6):581-586. [doi: 10.1016/j.jpsychores.2008.07.002] [Medline: 19027448]
44. Scheurer D, Choudhry N, Swanton KA, Matlin O, Shrank W. Association between different types of social support and medication adherence. Am J Manag Care 2012 Dec 01;18(12):e461-e467 [FREE Full text] [Medline: 23286676]
45. Kingod N, Cleal B, Wahlberg A, Husted GR. Online peer-to-peer communities in the daily lives of people with chronic illness: a qualitative systematic review. Qual Health Res 2017 Jan 12;27(1):89-99. [doi: 10.1177/1049732316680203] [Medline: 27956659]
46. van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar 01;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]
47. Stolldorf DP, Jones AB, Miller KF, Paz HH, Mumma BE, Danesh VC, et al. Medication discussions with patients with cardiovascular disease in the emergency department: an opportunity for emergency nurses to engage patients to support medication reconciliation. J Emerg Nurs 2023 Mar;49(2):275-286 [FREE Full text] [doi: 10.1016/j.jen.2022.12.002] [Medline: 36623969]
48. Serving Patients with Limited English Proficiency: Results of a Community Health Center Survey. National Association of Community Health Centers. 2013 Jul 23. URL: https://healthlaw.org/resource/serving-patients-with-limited-english-proficiency-results-of-a-community-healt/ [accessed 2024-12-26]

Abbreviations

FQHC: federally qualified health center
HCP: health care provider
PI: principal investigator
RCT: randomized controlled trial

Edited by A Kushniruk; submitted 25.06.24; peer-reviewed by P Cook, C Chiaranai; comments to author 11.09.24; revised version received 17.10.24; accepted 22.12.24; published 07.01.25.

Please cite as:
Hartch C, Dietrich MS, Lancaster BJ, Mulvaney SA, Stolldorf DP
Satisfaction and Usability of a Commercially Available Medication Adherence App (Medisafe) Among Medically Underserved Patients With Chronic Illnesses: Survey Study
JMIR Hum Factors 2025;12:e63653
URL: https://humanfactors.jmir.org/2025/1/e63653
doi:10.2196/63653
PMID:

Perception and Evaluation of a Knowledge Transfer Concept in a Digital Health Application for Patients With Heart Failure: Mixed Methods Study

Madeleine Flaucher[1], MSc; Sabrina Berzins[1], MSc; Katharina M Jaeger[1], MSc; Michael Nissen[1], MSc; Jana Rolny[2]; Patricia Trißler[2]; Sebastian Eckl[2]; Bjoern M Eskofier[1,3], Prof Dr; Heike Leutheuser[1], Dr. Ing.

[1]Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Carl-Thiersch-Str. 2b, Erlangen, Germany

[2]ProCarement GmbH, Forchheim, Germany

[3]Institute of AI for Health, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany

Corresponding Author:
Madeleine Flaucher, MSc
Machine Learning and Data Analytics Lab, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Carl-Thiersch-Str. 2b, Erlangen, Germany

Abstract

Background: Digital health education can enhance the quality of life of patients with heart failure by providing accessible and tailored information, which is essential for effective self-care and self-management.

Objective: This work aims to develop a mobile health knowledge transfer concept for heart failure in a user-centered design process grounded in theoretical frameworks. This approach centers on enhancing the usability, patient engagement, and meaningfulness of mobile health education in the context of heart failure.

Methods: A user-centered design process was employed. First, semistructured stakeholder interviews were conducted with patients (n=9) and medical experts (n=5). The results were used to develop a health knowledge transfer concept for a mobile health app for heart failure. This concept was implemented as a digital prototype based on an existing German mobile health app for patients with heart failure. We used this prototype to evaluate our concept with patients with heart failure in a study composed of user testing and semistructured patient interviews (n=7).

Results: Stakeholder interviews identified five themes relevant to mobile health education: individualization, content relevance, media diversity, motivation strategies, and trust-building mechanisms. The evaluation of our prototype showed that patients value the adaptation of content to individual interests and prior knowledge. Digital rewards such as badges and push notifications can increase motivation and engagement but should be used with care to avoid overload, irrelevance, and repetition.

Conclusions: Our findings emphasize the importance of tailoring mobile health education to the specific needs and preferences of patients with heart failure. At the same time, they also highlight the careful implementation of motivation strategies to promote user engagement effectively. These implications offer guidance for developing more impactful interventions to improve health outcomes for this population.

(JMIR Hum Factors 2025;12:e56798) doi:10.2196/56798

KEYWORDS

health literacy; digital Literacy; user-centered design; digital health app; heart failure; mixed methods study; user centered deign; usability; patient engagement; mHealth app; development

Introduction

In 2019, heart failure affected more than 56 million people worldwide [1]. The diagnosis of heart failure entails severe health consequences for patients and significant financial resources for public health systems [2]. In light of an aging population, heart failure is becoming increasingly important in society and requires new approaches for prevention, management, and therapy to mitigate health impact [3,4].

Heart failure requires a high level of active patient involvement [5]. Self-monitoring of physiological parameters, medication intake, nutrition, and physical activity are crucial factors in preventing the hospitalizations of patients [6,7]. In this context, health literacy plays a major role since high levels of health literacy in patients with heart failure promote patient empowerment, self-care practices, and medication adherence [8-10]. Health literacy can be described as the capacity of a person to access and understand health-related information, place it in context, and use it responsibly to promote and

maintain good health [11]. With a focus on heart failure and other cardiovascular diseases, The American Heart Association concludes that health literacy is crucial for effective treatments for cardiovascular diseases and acknowledges information technology as a potential path to improving health literacy [12]. In the long-term, low health literacy is associated with an increased risk for hospitalizations and death in patients with heart failure [13].

To effectively deliver information to patients and achieve learning effects, mobile health technology offers great potential. This applies to populations with low health literacy in particular [14-16]. As smartphone access has become widespread across various demographic groups, including those with lower socioeconomic status, they provide an accessible platform for delivering tailored health interventions to underserved populations [17,18]. Enhancing health literacy requires attention not only to disease-specific content but also to digital health literacy, empowering users to navigate health-related inquiries online effectively [14,19,20]. However, several barriers have been identified, including the access to apps, the readability of content as well as the usability of digital health services [14,21]. In the context of heart failure, evidence about the effects of mobile health for promoting health literacy is still limited. Allida et al identified no increase in heart failure knowledge through digital educational content and unclear evidence on self-efficacy, self-care, and health-related quality of life [22]. In contrast, several recent studies demonstrated positive effects on quality of life, hospitalization rates, and self-management abilities [23-25]. However, despite the potential benefits, these approaches lack the incorporation of theoretical frameworks for engaging, appealing, and effective knowledge transfer. Furthermore, patients with heart failure are hardly involved in the development process of mobile health interventions.

Thus, the aim of this work was to develop a concept for a sustainable and engaging knowledge transfer with a specific focus on the needs and preferences of patients with heart failure. Theoretical frameworks for mobile health knowledge transfer and the involvement of end-users throughout the design process build a strong foundation for an engaging, user-friendly, and effective mobile health intervention. A digital prototype served as a tangible representation of our concept, which we evaluated with patients with heart failure.

Methods

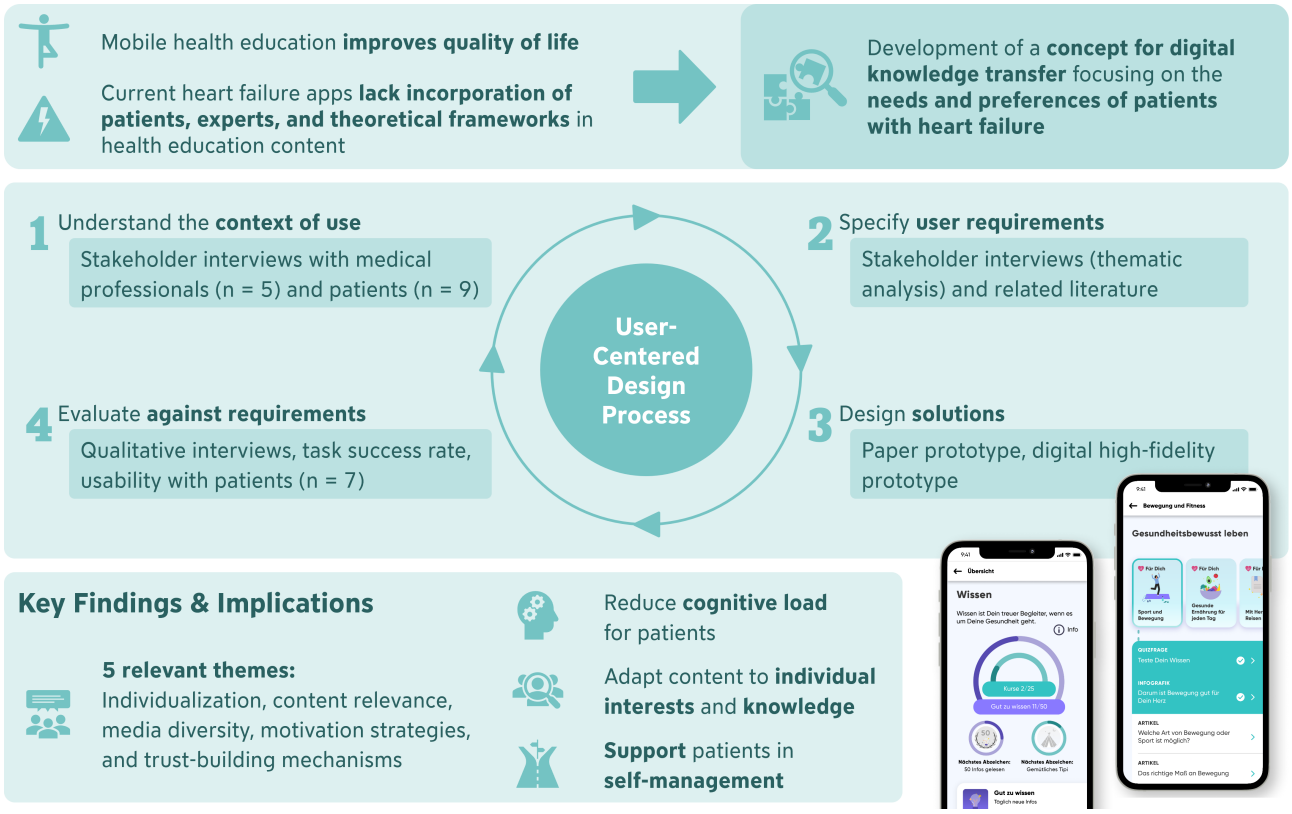
Study Design Overview

The foundation of this work was the user-centered design process, a common method in interaction design [26]. It emphasizes the involvement of the potential end-users throughout the developmental phases to ascertain that the resulting concepts align with their needs, preferences, and experiences of the target group. This iterative approach often incorporates techniques such as user interviews, focus groups, prototyping, and usability testing to refine the design based on continuous feedback. The graphical abstract (Figure 1) provides an overview of the methodology used.

First, we used current literature and semistructured qualitative interviews to understand the context of use and derive requirements. Based on the findings, possible design solutions were drafted as a paper prototype. Thereafter, a digital high-fidelity prototype was implemented. This digital prototype was then evaluated in a study focusing on the identified requirements of the target group.

The exemplary app we used for the prototype design is the ProHerz app (ProCarement Gmbh, Germany), a medical-grade mobile app for heart failure. This app is intended to support patients in managing their health by tracking vital signs and medication intake. Additionally, the company provides a ‘CareCenter,’ where, depending on the product used, medical experts regularly monitor the patients’ health parameters and stay in personal contact via messages and phone calls.

Figure 1. Graphical abstract illustrating the integration of the user-centered design process in the development of a health knowledge transfer concept for patients with heart failure.



Stakeholder Interviews

Recruitment

Qualitative semistructured interviews were conducted with patients (n=9) and medical experts (n=5). This approach is a flexible data collection method that combines predefined questions with the freedom to explore emerging topics to further enhance the understanding of participants' experiences and perspectives [27]. All patients were recruited from the ProHerz user base. Thus, all were patients with heart failure and previous experience with the ProHerz app. We chose this group because of their experience using a mobile health app for self-management. To allow for meaningful comparisons of feedback, our recruitment strategy aimed to include participants with varying levels of experience using the ProHerz app, aligning with best practices in end-user and usability testing. Medical experts were recruited through ProCarement, where they interact with the ProHerz users in their daily work. All experts were nurses with long-term experience in heart failure care. Their close exchange gives them a broad perspective on patients' characteristics, needs, and skills.

Procedure

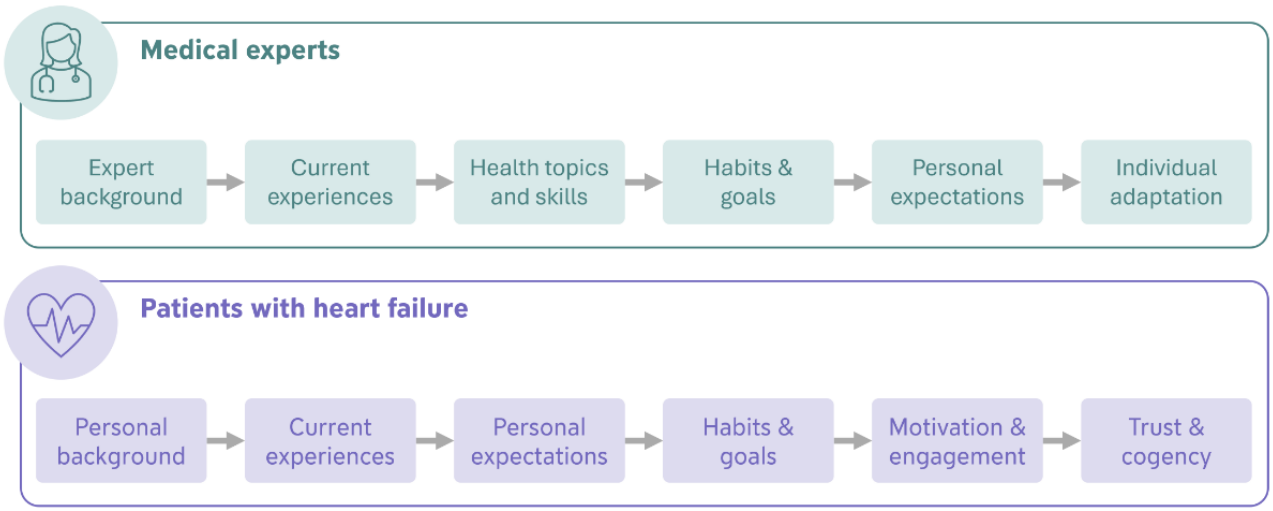
The interview guidelines were based on our research aims and current literature. Both stakeholder interviews covered similar aspects (Figure 2). The guidelines aimed to identify patients' specific needs and requirements regarding content and features within a mobile health knowledge transfer.

The interviews with the medical experts targeted explicitly at their perspective on essential knowledge and skill requirements that should be covered in a mobile health knowledge transfer concept. Besides, specific features, content presentation, and adaptation strategies for individual patient needs were discussed.

In the patient interviews, we inquired about experiences with specific features, their current interactions with the ProHerz app, and their expectations regarding knowledge transfer within a mobile health app designed for heart failure. Furthermore, the interviews included discussions on patients' health-related routines and goals and identifying requirements for trust-building and cogency. Additionally to the interview, patients were asked to complete the European Health Literacy Survey Questionnaire (EU-HLS-Q16) to assess their health literacy level [25,26]. Through this tool, participants rate general tasks related to health literacy based on their level of difficulty.

All interviews with medical experts were conducted via Microsoft Teams. Patient interviews were conducted in person.

Figure 2. The topics included in the interview with both stakeholder groups.



Prototype Development

The development of the knowledge transfer concept was based on the findings from the interviews and related literature.

We incorporated the framework of Riegel et al to structure the topics [28]. The content categories in the prototype are oriented toward the three self-care processes for heart failure: maintenance, symptom perception, and management. Information in the prototype was structured based on the principles for effective e-learning by Clark and Mayer and the recommendations for designing health-literate mobile apps by Broderick et al [29,30]. This included, for example, dividing information into manageable units, using different types of media, and writing in a conversational style.

Additionally, we considered possible cognitive limitations of users in the context of heart failure [22,31-34]. According to the Cognitive Load Theory, a reduction of cognitive load is able to improve learning outcomes. The theory describes three types of cognitive loads in the context of learning: intrinsic, extraneous, and germane cognitive load. We aimed at reducing the extraneous cognitive load by simplifying navigation, avoiding unnecessary elements, and ensuring a clear, uncluttered interface design. The intrinsic load was managed by sequencing information into smaller units aligned with the complexity of topics. The germane load was supported by including interactive elements such as quizzes or visual aids, encouraging active engagement.

Personalizing health education to address individual needs has been shown to improve health outcomes in areas such as physical activity, nutrition, and adherence to screening recommendations [35]. In the context of health education for heart failure, users face unique challenges through varying levels of health literacy and cognitive impairments [13]. As such, a personalized approach that takes relevant individual factors into account can meet diverse needs of patients, ensuring information is both accessible and relevant [22]. For this reason, we decided to incorporate the possibility of adapting the contents based on a short survey participants answered in the beginning.

The derived themes of the interviews were brought together with findings from the health literacy assessment that additionally revealed the needs of patients. We drafted a workflow for the mobile knowledge transfer and implemented this as a low-fidelity paper prototype.

For the first usability test, we embedded the paper prototype in Useberry (Useberry User Testing Technologies IKE, Greece). Useberry is a tool to test the usability of the paper-based prototype digitally. We aimed to find usability issues as early as possible in our design. For this initial testing, we invited healthy participants through personal contact. At this stage, the priority was gathering quick insights from users who were familiar with mobile apps to identify basic usability issues before moving on to the larger-scale user testing with patients with heart failure. This reduced the burden and required time and effort of the actual patients. Test users (n=5) were asked to perform different navigation-related tasks within the prototype. Outcome measures were the task completion rate and the recorded clicks.

After this initial usability testing, the paper prototype was then transferred into a digital high-fidelity prototype using Figma (Figma GmbH, Germany). This prototype included an example questionnaire for individualizing the information content based on the date of diagnosis, implanted devices (pacemaker and defibrillator), smoking habits, topics of interest, the preferred media formats and the New York Heart Association (NYHA) functional class (Figure 3). The NYHA classification is a commonly used system for categorizing the severity of heart failure based on physical activity limitations caused by cardiac symptoms. It ranges from Class I (no limitations) to Class IV (highest limitations) [36]. After the questionnaire, the basic landing page of the app was presented with an additional link to the health knowledge section. The main part of our prototype was the knowledge section with an overview and link to the available courses and the short information messages. Additionally, badges for learning-related achievements can be found there. Finally, an example course was implemented on the topic of sports and exercise (Figure 4).

Figure 3. Questionnaire for individual adaptation (left), home screen of the app with the link to the knowledge section (middle), and the main page of the knowledge section (right) of the digital prototype.

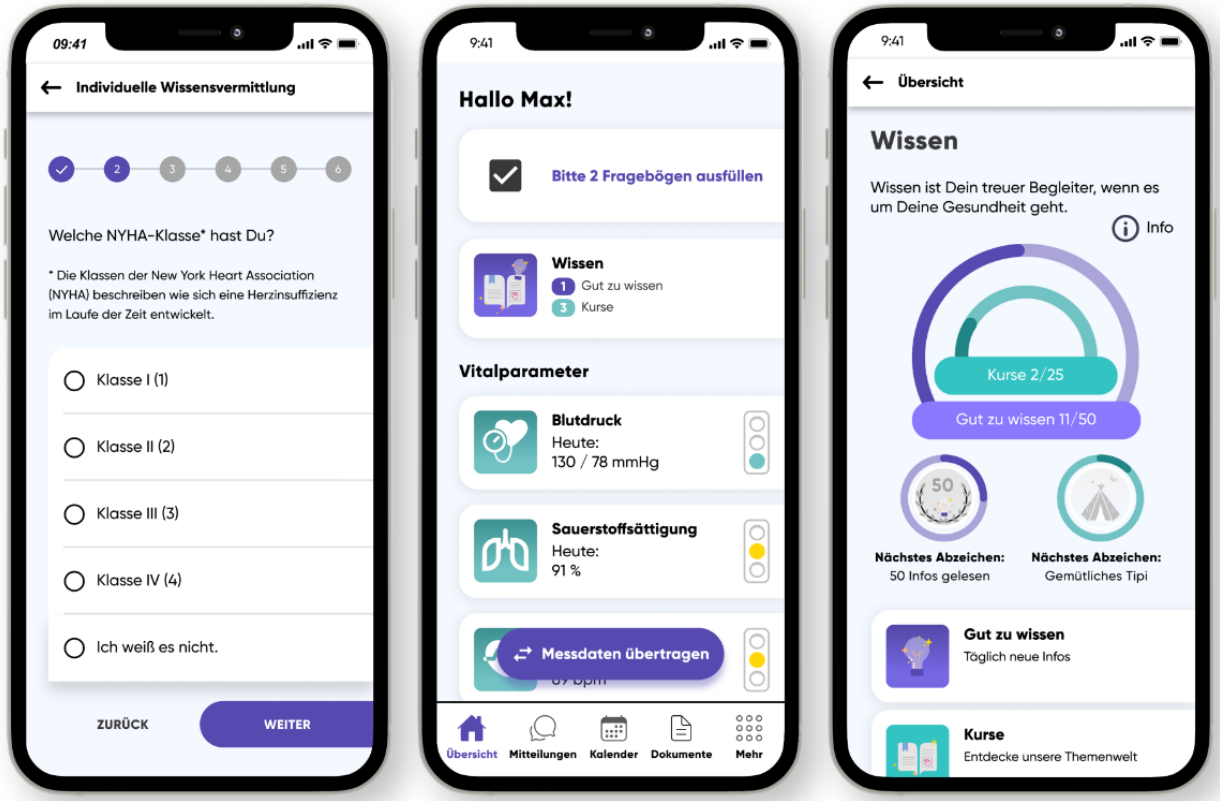
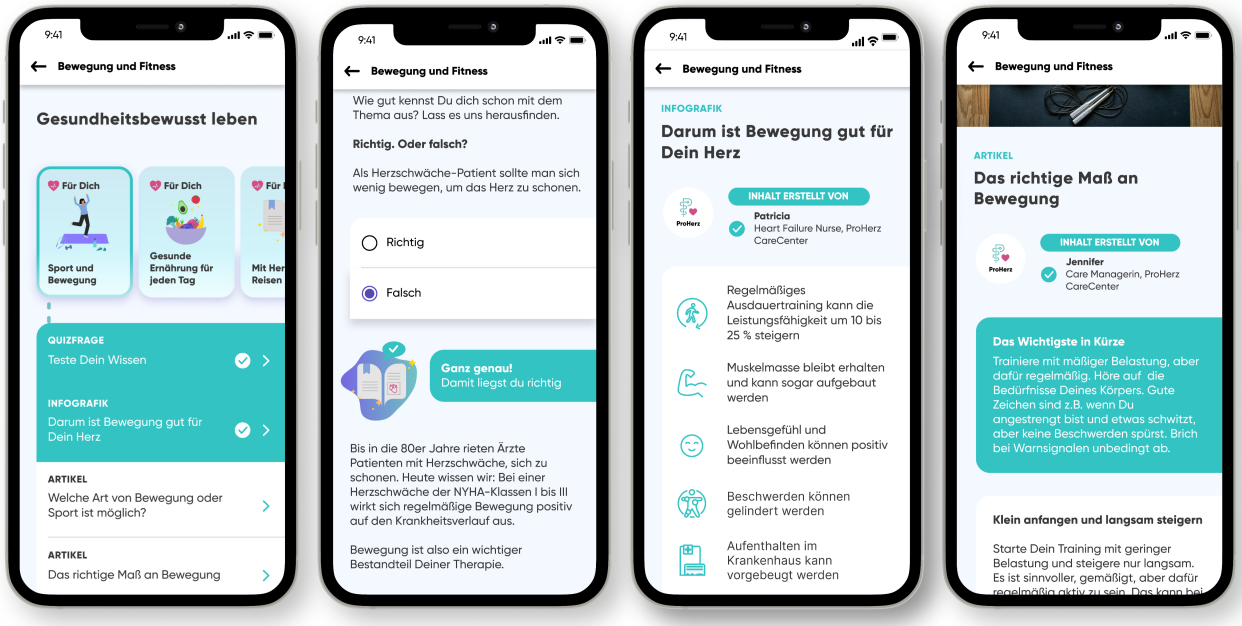


Figure 4. Structure of one exemplary course with (1) the overview of the content, (2) a quiz question, (3) an infographic presenting key facts, and (4) an article starting with information about the author of the text and a summary of the most important aspects of the article.



Concept Evaluation

Recruitment

For evaluating the knowledge transfer concept, we recruited patients with heart failure (n=7), who were already experienced in using a mobile health app. Therefore, all participants were recruited from the user base of ProCarement. Four participants already took part in the previous stakeholder interviews conducted as part of this study.

Procedure

After being instructed about the study procedure, participants were asked to complete 11 tasks with the digital prototype. Due to feasibility reasons, patients accessed the prototype on a computer interface. This approach allowed participants to interact with the app more comfortably during the evaluation, while also enabling researchers to observe user interactions in detail and collect qualitative feedback efficiently. Tasks included small, simple tasks and more complex, time-consuming tasks. An overview of all tasks is listed in Table 1. Afterwards, semistructured interviews were conducted to obtain the participants' overall impression and to identify features and characteristics that have the potential to improve engagement and increase health literacy.

Table . Tasks given to participants in the evaluation of the digital prototype of the health knowledge transfer concept.

No.	Description
1	Navigate to the questionnaire for individualization
2	Complete the questionnaire for individualization
3	Navigate to the knowledge section
4	Navigate to informative notifications, read the first unread notification
5	Navigate to the educational courses
6	Navigate to the course category 'symptoms'
7	Navigate to completed courses
8	Select a course within a course category
9	Navigate to recommended courses and select a recommended course
10	Complete one educational course
11	Navigate to the overview of badges

Analyses

All qualitative data obtained during interviews were analyzed using thematic analysis based on the approach of Braun and Clarke [37]. This method involves identifying, analyzing, and reporting patterns (themes) within the data, allowing for a rich and detailed interpretation of the participants' experiences and perspectives. The process includes familiarizing oneself with the data, generating initial codes, searching for themes, reviewing themes, and defining and naming them to ensure a comprehensive understanding of the qualitative insights.

Ethical Considerations

All presented substudies were approved by the institutional review board of the Friedrich-Alexander-Universität Erlangen-Nürnberg (approval no. 22 - 233-S), and we obtained written informed consent from all participants. The collected data were pseudonymized after collection to ensure participant confidentiality and data protection.

Results

Stakeholder Interviews

Demographics

The stakeholder interviews with patients and medical experts were conducted in August 2022. We recruited 9 patients with a mean (SD) age of 67 (7) years. Patients received their diagnosis between 15 and 3 years prior to this study. All patients were using the ProHerz app for at least 4 months (Table 2). Seven identified as women, and 2 identified as men. Additionally, 5 medical experts were recruited (Table 3) with work experience between 6 and 11 years.

Table . Characteristics of patients with heart failure included in the stakeholder interviews.

Interviewee ID	Age (y)	Identified as	Years since heart failure diagnosis	Months of app experience with ProHerz
P1	60	female	5	9
P2	58	female	3	5
P3	75	female	3	15
P4	78	female	5	15
P5	63	female	4	16
P6	75	female	9	4
P7	69	female	15	5
P8	60	male	3	4
P9	68	male	9	6

Table . Characteristics of medical experts included in the stakeholder interviews.

Interviewee ID	Identified as	Expert experience	Years of work experience with patients with heart failure
Expert 1	female	Heart failure nurse, licensed practical nurse	11
Expert 2	male	Licensed practical nurse	6
Expert 3	female	Licensed practical nurse	6
Expert 4	female	Bachelor of Science in Nursing, specialist in anesthesia and intensive care	9
Expert 5	male	State-qualified nurse	8

Thematic Analysis

We derived five themes relevant to a knowledge transfer concept from the stakeholder interviews with patients with heart failure and medical experts.

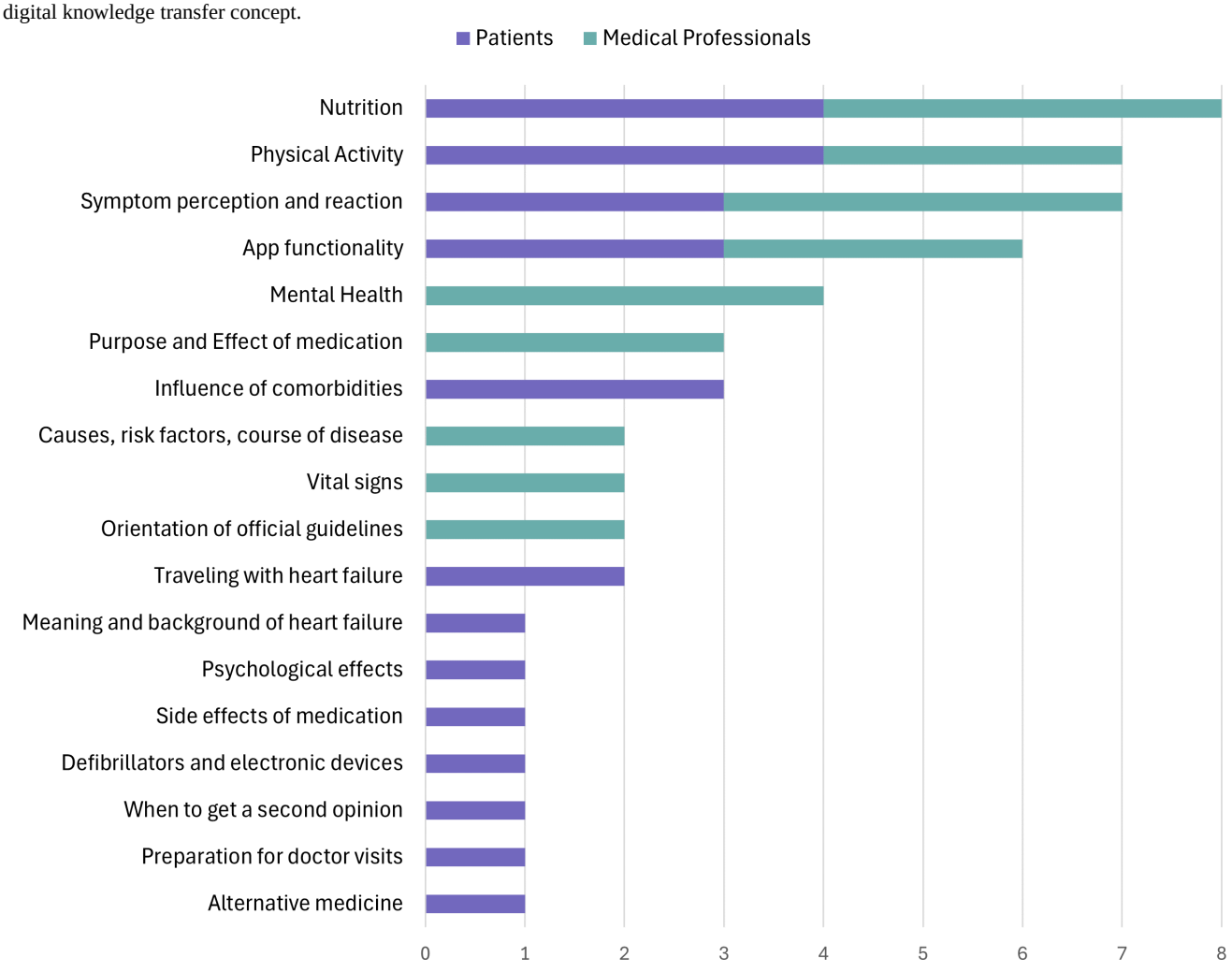
Individual Adaptation of Health Information

The interviews revealed a positive perspective towards the adaptation of content and push notifications to the different levels of knowledge of the users (4/5 experts, 2/9 patients). Our results suggest individualization based on different information such as the NYHA class, walking distance, app usage duration, physical condition age, prior knowledge, or personal preferences that could be determined for example through a quiz within the app.

Relevant Health Topics for Patients With Heart Failure

Several relevant health topics were identified within the stakeholder interviews (Figure 5). Health topics most requested by both patients and medical experts included nutrition (4/5 experts, 4/9 patients), physical activity (3/5 medical experts, 4/9 patients), symptom perception and reaction (4/5 medical experts, 3/9 patients), and app functionality (3/5 medical experts, 3/9 patients).

Figure 5. Health topics mentioned in the stakeholder interviews with patients and medical experts and that were considered relevant to be included in a digital knowledge transfer concept.



Diversity of Media Formats

Most of the interviewed patients (7/9 patients) preferred information in text format. For example, 1 participant stated, "you can read it two or three times if you say it's something that's difficult to absorb, maybe a text would be better" (P5, translated). Additionally, videos were perceived as useful by all medical experts but only desired by 2 patients. Related to the presentation of the information, patients mentioned using fewer technical terms and more simple language (2/9 patients) and shortly highlighting the most important aspects (2/9 patients). This can support easy comprehension of contents, which was also mentioned as a factor for the trustworthiness of health information (3/9 patients).

Promotion of Motivation, Engagement, and Joy

Medical experts (4/5 experts) suggested a reward system to increase engagement with the information contents. However, it should be ensured that this reward system is suitable for the specific target group of users (2/5). Three patients also mentioned the wish to receive positive feedback for progress within the app. Personal interest in the content was also mentioned (3/9) to have a positive influence on engagement.

Credibility and Trust

From the interviews with patients, we identified several factors related to the credibility of information. Positively perceived was the specification of authors (3/9 patients), especially if the authors are physicians (2/9 patients) or authors with an academic background (2/9 patients). In addition, the provider of the mHealth app should have a good reputation (3/9 patients). Factors restraining trust and credibility could be advertising (1/9 patients), authors without medical background (1/9 patients), or irrelevant information (1/9 patients).

Health Literacy

Health literacy levels based on the EU-HLS-Q16 were between 10 and 16 with an average (SD) score of 13.6 (2.1). Participants mostly had difficulties deciding on preventive actions based on media information (5/9 patients) and determining whether information about health risks in media is trustworthy (4/9 patients).

Evaluation of the Digital Prototype

Demographics

We evaluated the proposed digital prototype of the knowledge transfer concept with 7 patients with heart failure (Table 4). The mean (SD) age was 65 (6) years. Patients received their diagnosis between 2 and 24 years prior to this study. All patients were using the ProHerz app for at least 5 months. Four participants identified as women, and 3 as men. This group was less diverse in terms of ProHerz experience, which was a result of recruiting participants based on availability and willingness to participate.

Table . Characteristics of patients included in the evaluation study of the knowledge transfer concept.

Interviewee ID	Age (y)	Gender	Years since heart failure diagnosis	Months of app experience
P1	60	female	5	9
P2	58	female	3	5
P3	75	female	3	15
P5	63	female	4	16
P10	59	male	9	15
P11	71	male	2	15
P12	69	male	24	10

Usability

The mean task completion rate was 84%; 5 of the 11 tasks given during the study were completed by all participants. The lowest completion rate occurred with task 4, which was completed by 4 of 7 participants (Table 5).

Table . Task completion rate for all tasks performed by the participants within the usability testing.

No.	Description	Completion rate (%)
1	Navigate to the questionnaire for individualization	100
2	Complete the questionnaire for individualization	100
3	Navigate to the knowledge section	100
4	Navigate to informative notifications, read the first unread notification	57
5	Navigate to the educational courses	71
6	Navigate to the course category ‘symptoms’	71
7	Navigate to completed courses	71
8	Select a course within a course category	71
9	Navigate to recommended courses and select a recommended course	86
10	Complete one educational course	100
11	Navigate to the overview of badges	100

Interview Results

The thematic analysis of the semistructured interviews with patients within the evaluation of our proposed concept resulted in 3 main themes.

Knowledge Transfer

In the evaluation of the digital prototype, the majority of participants (4/7 patients) classified the prospective range of information as valuable. The organization and structure of contents received positive feedback from most participants (5/7 patients). Specifically, the clear arrangement of information (3/7 patients) and the definition of specific categories (2/7 patients) were highlighted. Patients also perceived the individual adaptation as positive, because personally irrelevant content would be hidden (4/7 patients).

To further improve the concept, participants indicated the necessity to update contents in response to potentially evolving needs over time due to lifestyle or health condition changes (2/7 patients).

Motivation and Engagement

A majority (5/7 patients) expressed an intention to utilize the course content in the future. Notably, reminders about feasible health-promoting measures were identified as a motivational factor by 3 out of 7 patients, with a specific emphasis from 2 participants on messages related to physical activity.

Concerns were raised about the potential negative impact of irrelevant, repetitive, or overly frequent notifications (4/7 patients). Conversely, aligning information with individual interests was identified as a key factor in promoting joy and sustaining engagement among users.

Patients expressed joy while answering the quiz questions, which may have contributed to a more engaging experience (2/7 patients). There was a suggestion to place quiz questions at the end of a course to assess the comprehension of the content effectively.

Participants generally responded positively to medical staff monitoring of progress, considering it a supportive element in their engagement with the course. However, 1 participant viewed social comparison negatively. Furthermore, intangible rewards, such as badges, were not universally well-received, with 1 out of 7 participants expressing a negative opinion, stating that it is solely virtual and not a real reward (P11).

Ease of Use

Out of 7 patients, 3 patients explicitly mentioned the app as easy to use. Other participants did not comment on this aspect, neither positively nor negatively. Concerns regarding the ease of use included potential initial overload when first using the app and potential individual difficulties with the technology and navigation.

Discussion

Principal Findings and Implications

In this study, we developed a concept for mobile health knowledge transfer for patients with heart failure. We grounded this concept on the requirements of patients with heart failure and current theoretical frameworks. From the patient and caregiver perspectives, we derived five central themes for a health knowledge transfer concept.

Our findings contribute to the understanding of user preferences and content organization in digital health interventions targeted at users with heart failure. The integration of diverse methodologies, including qualitative interviews, usability testing, and participant feedback, fortifies the study’s comprehensiveness, providing insights into user engagement, health literacy promotion, and the practical implications for digital health interventions. The results of this study represent an important step toward the development of a mobile health knowledge transfer tool for heart failure patients. By identifying relevant features and characteristics, this work provides a foundation for further refinement and testing, ultimately aiming to support health literacy and improve patient outcomes. It

becomes evident that a concept should focus on individualization, topic relevance, media diversity, motivation strategies, and trust-building mechanisms to optimize its impact on health literacy and engagement of patients with heart failure. Based on the findings, identified in the stakeholder interviews, we developed a concept for knowledge transfer in a digital health app for patients with heart failure.

We evaluated our concept with a digital prototype and identified features and characteristics that are perceived as supportive in increasing health literacy, that can promote motivation, engagement, and joy, and that are able to convince patients of a mobile health knowledge transfer. Most prominent was the wish for content that is individually adapted and balanced regarding the topics covered as well as media formats used to present those topics.

The results of the prototype evaluation presented both encouraging feedback and areas for refinement. Participants highlighted the app's structured design and ease of navigation, emphasizing that these features facilitated a straightforward user experience and enhanced the clarity of the presented information. The quiz elements, in particular, were frequently mentioned as engaging and motivating, showcasing the potential of interactive components to sustain user interest and reinforce learning. These positive responses suggest that the app effectively addressed some of its primary goals, including improving knowledge transfer and supporting patient engagement.

Conversely, the feedback also identified challenges that should be addressed in future iterations. Participants expressed concerns of getting overwhelmed, potentially hindering comprehension. This may reflect concerns about managing excessive information or complex interactions within the app, particularly in the context of their health condition. For users with chronic conditions like heart failure, the introduction of too much content or functionality in a single interface could exacerbate feelings of being overwhelmed and lead to disengagement or frustration [38]. Furthermore, badges and achievements were mentioned to be not motivating despite research showing the positive effects of such gamification elements [39,40]. Older users or those less familiar with gamification might find such features irrelevant or even patronizing if they do not see clear value in them. Other gamification elements such as progress bars, interactive storylines, or more useful rewards may be more effective in motivating users. Therefore, future iterations could explore the inclusion of a simple setup and customization menu to enable or disable features such as gamification and other user-preferred functionalities.

Accessibility is a critical consideration in designing digital health systems to ensure they meet the needs of diverse users. While none of the participants in our study raised concerns or reported difficulties with accessibility, indicating that the system was well-received by the target population, it is important to account for potential challenges in broader user groups. Features such as larger fonts, high-contrast visuals, simplified navigation, and optional audio support can further enhance usability, particularly for individuals with age-related changes in vision, hearing, or varying levels of digital literacy.

The high rates of completion for most tasks suggest that the interface and navigation design are intuitive, enabling users to locate and interact with the desired functions efficiently. This aligns with the positive feedback received during the interviews conducted within the evaluation. However, with a few tasks having a lower completion rate, some aspects of the design need further refinement in future iterations. Factors such as differences in technology familiarity, cognitive load, or previous experience with the ProHerz app could have contributed to these discrepancies. Interactive tutorials to onboard users and a further adaptation of content may be able to better account for different requirements, levels of experience, or cognitive abilities. In particular, task 4 with a completion rate of only 57% was the most problematic task for participants. The participants were not able to locate the button to navigate to the informative notifications. The reason for this is most likely the design of the progress bar on the same page (Figure 3, right) that users expected to be the button to the next page. Ensuring more consistent visual cues for interactive elements could further reduce the potential for misinterpretation and enhance the overall user experience.

Comparison With Prior Work

The need for individualization is in line with the findings of Giordan et al [41], calling for further customization especially related to the educational level and digital literacy of users. Our results additionally highlight the need for customization regarding personal interests and lifestyle and the importance of regular adaptation to changing circumstances in the life and knowledge of users. Future applications require new strategies to individually adapt educational content that can change over time. This can include regularly soliciting user feedback through surveys, using quizzes to assess understanding, or adaptive learning algorithms based on user behavior within the mobile app [42].

The topics covered in our knowledge transfer concept are in line with the recommendations of the European Society of Cardiology. They included information about heart failure, symptom monitoring, self-management, medication, fluid intake, implantable devices, physical activity, nutrition, smoking, sleep mental health, and traveling [43]. The information was presented using different media formats, which relates to current literature emphasizing the benefits of this approach. A variety of formats can support users with different health literacy levels and can deepen the learning effects [29,43,44].

Strengths and Weaknesses

Several limitations apply to this study. Firstly, the participant pool featured a restricted number of individuals, all German-speaking and residents in Germany. Moreover, participants were well-acquainted with the ProHerz app, potentially influencing their perspectives and feedback, especially regarding the ease of use. Given the advanced health literacy levels of the participants, our findings may not comprehensively represent the needs of individuals with lower health literacy. The initial usability testing was conducted with a convenience sample due to limited patient availability; therefore, results from this group need to be interpreted with caution. During the evaluation phase, 4 participants already

were part of the stakeholder interviews in the first phase of this study due to limited participant availability. Lastly, all participants accessed the app on a laptop rather than a mobile phone, introducing a potential discrepancy, as some users reported difficulties with computer usage and the overall interaction experience may be different. These limitations underscore the necessity for cautious interpretation and highlight avenues for future research refinement.

Our study combines several strengths: we were the first to conceptualize a mobile health knowledge transfer within a user-centered design approach. By prioritizing patient needs and requirements, and involving healthcare professionals to ensure relevance and meaningfulness, the study sets a precedent in fostering impactful and patient-centric digital health interventions. The incorporation of theoretical frameworks for mobile health education enriches the conceptual foundation, contributing to the overall robustness of the study and its potential impact on the enhancement of patient-centered care.

Conclusions

We were the first to develop a mobile health knowledge transfer for heart failure grounded on theoretical frameworks within a user-centered design process.

By addressing individual health literacy levels, cognitive challenges, and personal preferences, this knowledge transfer concept can help patients better understand and manage their condition. The active involvement of healthcare professionals ensures that the resulting intervention is meaningful and aligned with relevant healthcare aspects. The findings also emphasize a discrepancy between what healthcare professionals perceive as important for patients and what patients themselves prioritize, particularly when it comes to identifying relevant healthcare topics. Our results are paving the way to more personalized and effective mobile health education for individuals managing heart failure. The concept can be used as a foundation for other digital platforms and mobile health apps to provide tailored educational content and supporting clinicians in delivering consistent, patient-specific information in home settings.

Future research opportunities should focus on longitudinal studies to assess the sustained impact of this knowledge transfer concept on health literacy, self-care behavior, and patient outcomes. In this context, validated usability scales alongside user interviews should be incorporated during iterative development processes, while also addressing potential biases in design and testing to ensure more robust and generalizable findings. Additionally, possible strategies for better individualization of health information should be further explored. To better understand the impact of personalization, future studies could incorporate a comparison of user experiences before and after customizing the app, providing valuable insights into how individualization influences usability and user satisfaction. This presents an opportunity to investigate adaptive learning algorithms that dynamically tailor content based on user progress and preferences.

Acknowledgments

This project is supported by the Bavarian State Ministry of Health and Care, project grant number PBN-MGP-2010-0004-DigiOnko. The authors would like to acknowledge the financial support from Deutsche Forschungsgemeinschaft and Friedrich–Alexander Universität Erlangen-Nürnberg within the funding program Open Access Publication Funding.

Authors' Contributions

MF conceived the study, developed the study design, analyzed data, assisted in interviews, and wrote the manuscript draft. SB and JR assisted in developing the study design, interview conduction, and data analysis. KMJ, MN, and HL assisted in writing the manuscript. PT, SE, BME, and HL initiated and supervised the project. All authors reviewed the final manuscript.

Conflicts of Interest

JR, SE, and PT are employees of ProCarement GmbH. All other authors declare no conflicts of interest.

References

1. Yan T, Zhu S, Yin X, et al. Burden, trends, and inequalities of heart failure globally, 1990 to 2019: a secondary analysis based on the global burden of disease 2019 study. J Am Heart Assoc 2023 Mar 21;12(6):e027852. [doi: 10.1161/JAHA.122.027852] [Medline: 36892088]
2. Heidenreich PA, Fonarow GC, Opsha Y, et al. Economic issues in heart failure in the United States. J Card Fail 2022 Mar;28(3):453-466. [doi: 10.1016/j.cardfail.2021.12.017] [Medline: 35085762]
3. Roth GA, Forouzanfar MH, Moran AE, et al. Demographic and epidemiologic drivers of global cardiovascular mortality. N Engl J Med 2015 Apr 2;372(14):1333-1341. [doi: 10.1056/NEJMoa1406656] [Medline: 25830423]
4. Guasti L, Dilaveris P, Mamas MA, et al. Digital health in older adults for the prevention and management of cardiovascular diseases and frailty. A clinical consensus statement from the ESC Council for Cardiology Practice/Taskforce on Geriatric Cardiology, the ESC Digital Health Committee and the ESC Working Group on e-Cardiology. ESC Heart Fail 2022 Oct;9(5):2808-2822. [doi: 10.1002/ehf2.14022] [Medline: 35818770]
5. Dunn P, Margaritis V, Anderson C. Understanding health literacy skills in patients with cardiovascular disease and diabetes. TQR 2017 Jan 2. [doi: 10.46743/2160-3715/2017.2548]

6. Michalsen A, König G, Thimme W. Preventable causative factors leading to hospital admission with decompensated heart failure. Heart 1998 Nov;80(5):437-441. [doi: 10.1136/hrt.80.5.437] [Medline: 9930040]
7. Wen CP, Wai JPM, Tsai MK, et al. Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study. The Lancet 2011 Oct;378(9798):1244-1253. [doi: 10.1016/S0140-6736(11)60749-6]
8. Evangelista LS, Rasmusson KD, Laramee AS, et al. Health literacy and the patient with heart failure--implications for patient care and research: A consensus statement of the Heart Failure Society of America. J Card Fail 2010 Jan;16(1):9-16. [doi: 10.1016/j.cardfail.2009.10.026] [Medline: 20123313]
9. Matsuoka S, Tsuchihashi-Makaya M, Kayane T, et al. Health literacy is independently associated with self-care behavior in patients with heart failure. Patient Educ Couns 2016 Jun;99(6):1026-1032. [doi: 10.1016/j.pec.2016.01.003] [Medline: 26830514]
10. Wu JR, Reilly CM, Holland J, Higgins M, Clark PC, Dunbar SB. Relationship of health literacy of heart failure patients and their family members on heart failure knowledge and self-care. J Fam Nurs 2017 Feb;23(1):116-137. [doi: 10.1177/1074840716684808] [Medline: 28795936]
11. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12(1):80. [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
12. Magnani JW, Mujahid MS, Aronow HD, et al. Health literacy and cardiovascular disease: fundamental relevance to primary and secondary prevention: a scientific statement from the American Heart Association. Circulation 2018 Jul 10;138(2):e48-e74. [doi: 10.1161/CIR.0000000000000579] [Medline: 29866648]
13. Fabbri M, Murad MH, Wennberg AM, et al. Health Literacy and outcomes among patients with heart failure: a systematic review and meta-analysis. JACC Heart Fail 2020 Jun;8(6):451-460. [doi: 10.1016/j.jchf.2019.11.007] [Medline: 32466837]
14. Kim H, Xie B. Health literacy in the eHealth era: A systematic review of the literature. Patient Educ Couns 2017 Jun;100(6):1073-1082. [doi: 10.1016/j.pec.2017.01.015] [Medline: 28174067]
15. Conard S. Best practices in digital health literacy. Int J Cardiol 2019 Oct 1;292:277-279. [doi: 10.1016/j.ijcard.2019.05.070] [Medline: 31230937]
16. Zippel-Schultz B, Palant A, Helms TM. Digitale Lösungen für die versorgung von herzinsuffizienzpatienten. In: Pfannstiel MA, Holl F, Swoboda WJ, editors. mHealth-Anwendungen Für Chronisch Kranke Wiesbaden [Book in German]: Springer Fachmedien Wiesbaden; 2020:135-155. [doi: 10.1007/978-3-658-29133-4_8]
17. Beneito-Montagut R, Rosales A, Fernández-Ardèvol M. Emerging digital inequalities: a comparative study of older adults' smartphone use. Social Media + Society 2022 Oct;8(4):20563051221138756. [doi: 10.1177/20563051221138756]
18. Mackert M, Champlin SE, Holton A, Muñoz II, Damásio MJ. eHealth and Health Literacy: a research methodology review. J Comput-Mediat Comm 2014 Apr;19(3):516-528. [doi: 10.1111/jcc4.12044]
19. Mein E, Fuentes B, Soto Más F, Muro A. Incorporating digital health literacy into adult ESL education on the US-Mexico border. Rhetor Prof Commun Glob 2012 Dec;3(1):162-174. [Medline: 23730533]
20. Blunck D, Kastner L, Nissen M, Winkler J. The effectiveness of patient training in inflammatory bowel disease knowledge via Instagram: randomized controlled trial. J Med Internet Res 2022 Oct 19;24(10):e36767. [doi: 10.2196/36767] [Medline: 36260385]
21. Klimis H, Thakkar J, Chow CK. Breaking barriers: mobile health interventions for cardiovascular disease. Can J Cardiol 2018 Jul;34(7):905-913. [doi: 10.1016/j.cjca.2018.02.012] [Medline: 29887218]
22. Allida S, Du H, Xu X, et al. mHealth education interventions in heart failure. Cochrane Database Syst Rev 2020 Jul 2;7(7):CD011845. [doi: 10.1002/14651858.CD011845.pub2] [Medline: 32613635]
23. Bakogiannis C, Tsarouchas A, Mouselimis D, et al. A patient-oriented app (ThessHF) to improve self-care quality in heart failure: from evidence-based design to pilot study. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e24271. [doi: 10.2196/24271] [Medline: 33847599]
24. Ahn JA. Heart failure-smart life ii: Effects of a mobile-based education and self-management program for patients with heart failure and their family caregivers. J Am Coll Cardiol 2022 Mar;79(9):2027. [doi: 10.1016/S0735-1097(22)03018-2]
25. Heiney SP, Donevant SB, Arp Adams S, Parker PD, Chen H, Levkoff S. A smartphone app for self-management of heart failure in older African Americans: feasibility and usability study. JMIR Aging 2020 Apr 3;3(1):e17142. [doi: 10.2196/17142] [Medline: 32242822]
26. Richter M, Flückiger M. User-Centred Engineering: Springer Berlin Heidelberg; 2014. [doi: 10.1007/978-3-662-43989-0]
27. Adeoye‐Olatunde OA, Olenik NL. Research and scholarly methods: semi‐structured interviews. J Am Coll Clin Pharm 2021 Oct;4(10):1358-1367 [FREE Full text] [doi: 10.1002/jac5.1441]
28. Riegel B, Dickson VV, Faulkner KM. The situation-specific theory of heart failure self-care. J Cardiovasc Nurs 2016;31(3):226-235. [doi: 10.1097/JCN.0000000000000244] [Medline: 25774844]
29. Clark RC, Mayer RE. E-Learning and the Science of Instruction: Proven Guidelines for Consumers and Designers of Multimedia Learning, 4th edition: Hoboken: Wiley; 2016.
30. Broderick J, Devine T, Lemerise AJ, et al. Designing health literate mobile apps. NAM Perspectives 2014 Jan 28;4(1) [FREE Full text] [doi: 10.31478/201401a]
31. van den Hurk K, Reijmer YD, van den Berg E, et al. Heart failure and cognitive function in the general population: the Hoorn Study. Eur J Heart Fail 2011 Dec;13(12):1362-1369. [doi: 10.1093/eurjhf/hfr138]

32. Currie K, Rideout A, Lindsay G, Harkness K. The association between mild cognitive impairment and self-care in adults with chronic heart failure: a systematic review and narrative synthesis. J Cardiovasc Nurs 2015;30(5):382-393. [doi: 10.1097/JCN.0000000000000173] [Medline: 24988321]
33. Vogels RLC, Scheltens P, Schroeder‐Tanka JM, Weinstein HC. Cognitive impairment in heart failure: a systematic review of the literature. European J of Heart Fail 2007 May;9(5):440-449. [doi: 10.1016/j.ejheart.2006.11.001]
34. Wilson EAH, Wolf MS. Working memory and the design of health materials: a cognitive factors perspective. Patient Educ Couns 2009 Mar;74(3):318-322. [doi: 10.1016/j.pec.2008.11.005] [Medline: 19121915]
35. Cohn WF, Lyman J, Broshek DK, et al. Tailored educational approaches for consumer health: a model to address health promotion in an era of personalized medicine. Am J Health Promot 2018 Jan;32(1):188-197. [doi: 10.1177/0890117116671082] [Medline: 29214832]
36. Bennett JA, Riegel B, Bittner V, Nichols J. Validity and reliability of the NYHA classes for measuring research outcomes in patients with cardiac disease. Heart Lung 2002 Jul;31(4):262-270. [doi: 10.1067/mhl.2002.124554]
37. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
38. Issom DZ, Henriksen A, Woldaregay AZ, Rochat J, Lovis C, Hartvigsen G. Factors influencing motivation and engagement in mobile health among patients with sickle cell disease in low-prevalence, high-income countries: qualitative exploration of patient requirements. JMIR Hum Factors 2020 Mar 24;7(1):e14599. [doi: 10.2196/14599] [Medline: 32207692]
39. Shakoor A, Mohansingh C, van der Boon RMA, Brugts JJ, Schaap J. Gamification and its potential for better engagement in the management of heart failure or quality of care registries: a viewpoint. Curr Heart Fail Rep 2025 Dec;22(1):4. [doi: 10.1007/s11897-024-00692-8]
40. Berglund A, Jaarsma T, Berglund E, Strömberg A, Klompstra L. Understanding and assessing gamification in digital healthcare interventions for patients with cardiovascular disease. Eur J Cardiovasc Nurs 2022 Aug 29;21(6):630-638. [doi: 10.1093/eurjcn/zvac048] [Medline: 35709297]
41. Bezerra Giordan L, Tong HL, Atherton JJ, et al. The use of mobile apps for heart failure self-management: systematic review of experimental and qualitative studies. JMIR Cardio 2022 Mar 31;6(1):e33839. [doi: 10.2196/33839] [Medline: 35357311]
42. Tong HL, Quiroz JC, Kocaballi AB, et al. Personalized mobile technologies for lifestyle behavior change: A systematic review, meta-analysis, and meta-regression. Prev Med 2021 Jul;148:106532. [doi: 10.1016/j.ypmed.2021.106532] [Medline: 33774008]
43. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. Eur Heart J 2021 Sep 21;42(36):3599-3726. [doi: 10.1093/eurheartj/ehab368] [Medline: 34447992]
44. Smith B, Magnani JW. New technologies, new disparities: The intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 1;292:280-282. [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]

Abbreviations

NYHA: New York Heart Association

Edited by A Kushniruk; submitted 05.04.24; peer-reviewed by R Marshall, W Luan; revised version received 24.01.25; accepted 28.01.25; published 31.03.25.

Please cite as:
Flaucher M, Berzins S, Jaeger KM, Nissen M, Rolny J, Trißler P, Eckl S, Eskofier BM, Leutheuser H
Perception and Evaluation of a Knowledge Transfer Concept in a Digital Health Application for Patients With Heart Failure: Mixed Methods Study
JMIR Hum Factors 2025;12:e56798
URL: https://humanfactors.jmir.org/2025/1/e56798
doi:10.2196/56798

The Utilization of Electronic Consultations (eConsults) to Address Emerging Questions Related to Long COVID-19 in Ontario, Canada: Mixed Methods Analysis

Jatinderpreet Singh[1,2], MD, MSc, MASc; Michael Quon[3,4,5], MD; Danica Goulet[6,7], MSc; Erin Keely[5,7,8,9,10], MD; Clare Liddy[2,6,7,9,10], MD, MSc

[1]Department of Family Medicine, Queen's University, Kingston, ON, Canada
[10]Ontario Health, Toronto, ON, Canada
[2]Department of Family Medicine, University of Ottawa, Ottawa, ON, Canada
[3]Department of Medicine (General Internal Medicine), The Ottawa Hospital, Ottawa, ON, Canada
[4]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[5]Department of Medicine, University of Ottawa, Ottawa, ON, Canada
[6]C T Lamont Primary Health Care Research Centre, Bruyère Health Research Institute, Ottawa, ON, Canada
[7](Ontario) eConsult Centre of Excellence, The Ottawa Hospital, Ottawa, ON, Canada
[8]Division of Endocrinology and Metabolism, The Ottawa Hospital, Ottawa, ON, Canada
[9]Ottawa Hospital Research Institute, Ottawa, ON, Canada

Corresponding Author:
Clare Liddy, MD, MSc
Department of Family Medicine, University of Ottawa, Ottawa, ON, Canada

Abstract

Background: Long COVID is an often debilitating condition affecting millions of people. Its diverse clinical presentations make effective diagnosis and management at the primary care level difficult, while specialist services for long COVID face extensive wait times. An electronic consultation (eConsult) program in Ontario developed a long COVID specialist group to allow primary care providers (PCPs) prompt access to specialist advice for patients with long COVID.

Objective: This study aims to assess patterns of service use, response times, impact, and clinical content of eConsult cases submitted to an eConsult long COVID specialist group in Ontario.

Methods: This study is a mixed methods analysis of eConsults submitted by PCPs to the long COVID specialist group of 2 eConsult services (Champlain eConsult BASE and Ontario eConsult) between June 1, 2021, and July 31, 2022. Data sources included the use data collected automatically by the services, responses to a mandatory closeout survey, and the content of PCP questions and specialist responses (Champlain eConsult BASE service only). Clinical questions or responses were analyzed using 2 validated taxonomies. Descriptive statistics were used for survey responses and use data.

Results: A total of 40 PCPs submitted 47 eConsults through Champlain eConsult BASE and 197 PCPs submitted 228 cases through Ontario eConsult. The median specialist response time was 0.6 (IQR 0.19-2.36; mean 1.7, SD 2.29) days. The 5 most common symptoms of long COVID were fatigue (14/47, 30%), dyspnea (7/47, 15%), cough (6/47, 13%), altered sense of smell (ie, anosmia and parosmia; 6/47, 13%), and cognitive changes (6/47, 13%). The five main question categories asked by PCPs were: (1) management of chronic symptoms of COVID-19, (2) need for additional work-up or follow-up testing, (3) community resources to support or manage patients with long COVID, (4) diagnostic clarification, and (5) guidance regarding COVID-19 vaccination.

Conclusions: The long COVID groups provided rapid access to a multispecialty service that facilitated the avoidance of unnecessary face-to-face referrals. An assessment of eConsults highlighted 5 common question types, providing insight into potential gaps in knowledge among PCPs that could help guide medical education and policy.

(JMIR Hum Factors 2025;12:e58582) doi:10.2196/58582

KEYWORDS

COVID-19; long COVID; eConsult; consultation; Canada; mixed methods analysis; diagnosis; primary care; electronic consultation; COVID specialist; specialist; patient; assessment; COVID-19 vaccination; vaccination; symptom; medical education; web-based consultation; teleconsultation; web-based consultations

Introduction

The COVID-19 pandemic has had an unprecedented impact on health care systems and economies worldwide. As of November 2023, there have been over 700 million reported cases of COVID-19, resulting in over 6.5 million deaths [1]. As the pandemic began spreading, the potential for long-term morbidity from COVID-19 became apparent. The World Health Organization reports that 10% - 20% of those infected with COVID-19 report incomplete recovery months beyond the acute illness—a condition commonly referred to as long COVID or post-COVID-19 condition [2].

A recent Statistics Canada survey reported that 14.8% of respondents with a previous positive test or suspected infection for COVID-19 experienced ongoing symptoms at least 3 months after infection [3]. Potential symptoms of long COVID are varied, and people often experience symptoms across multiple organ systems.

Primary care is ideally positioned to diagnose and provide management for patients with long COVID. Unfortunately, given its diverse clinical presentations and limited empirical evidence guiding management, many primary care providers (PCPs) lack the specialized knowledge to effectively identify and manage long COVID. In certain jurisdictions, long COVID specialty clinics have been established to support primary care. However, demand and wait times for these clinics make timely access a challenge. There is a need for innovative tools to support PCPs in identifying and delivering care for those with long COVID.

One solution to these challenges is electronic consultation (eConsult), which allows PCPs to ask clinical questions regarding their patients to specialists using a secure web-based platform. In response to the pandemic, an eConsult service operating in Ontario, Canada, launched a long COVID specialist group, allowing PCPs to connect with specialists across multiple disciplines with expertise in long COVID to help with the diagnosis and management of their patients. Previous studies assessing this eConsult service in Ontario have demonstrated improved access to specialty care, reduced need for face-to-face specialist visits, cost savings, and high physician satisfaction [4-8].

In this study, we examined eConsult cases submitted to a long COVID specialist group in Ontario to assess patterns of service use, response times, the impact on the need for formal face-to-face referrals, and the content of clinical questions being asked.

Methods

Study Design

This study is a mixed methods analysis of eConsults submitted by PCPs to the long COVID specialist group between June 1, 2021, and July 31, 2022.

Ethical Considerations

This study received ethics approval from the Ottawa Health Science Network Research Ethics Board (REB protocol 2009848 - 01). The Ottawa Health Science Network Research Ethics Board waived informed consent for the study due to its retrospective, cross-sectional nature, so individualized patient consent was not obtained. All nonanonymized data were stored on secure servers that are accessible only by approved users who signed a privacy and confidentiality form.

Participants

Participants in the study were PCPs who were registered to 1 of the 2 eConsult services and who submitted at least 1 eConsult to the long COVID group within the study timeframe.

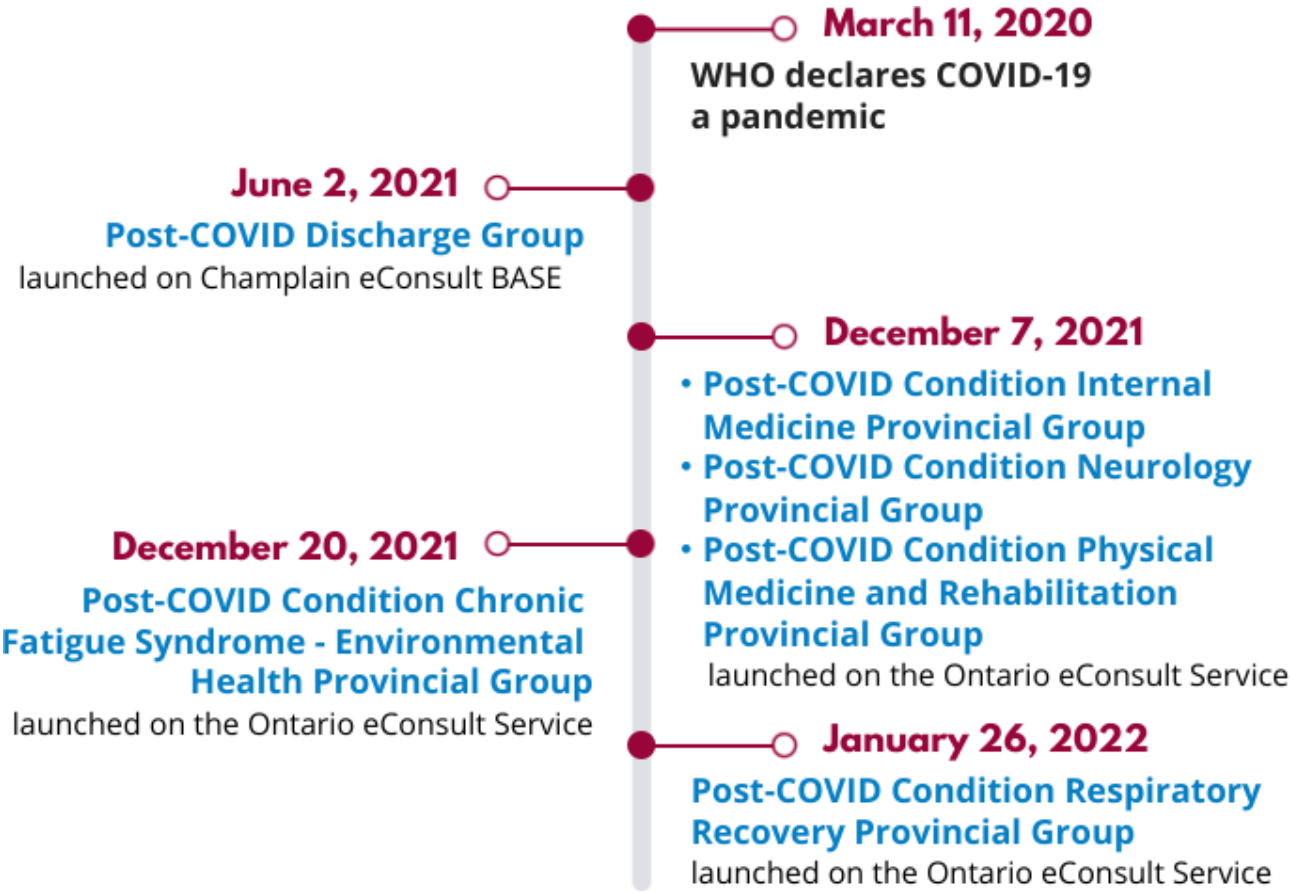
eConsult Services

Two multispecialty eConsult services currently operate in Ontario under the umbrella of Ontario eConsult. The first service, the Champlain eConsult BASE (Building Access to Specialists through eConsultation) service, was initially piloted in 2010 across the Champlain region of eastern Ontario, which includes Ottawa and its surrounding communities and has a population of around 1.3 million people. The second service, the Ontario eConsult Service, made eConsult available to all PCPs across Ontario in 2018. Both services are funded through the Ontario Ministry of Health. On June 2, 2021, a long COVID specialty group was established on the Champlain eConsult BASE service. Later in the year, a similar service was made available on the Ontario eConsult Service, providing PCPs with access to five different long COVID specialty subgroups that were rolled out between December 2021 to January 2022: (1) Post-COVID Condition Internal Medicine Provincial Group, (2) Post-COVID Condition Neurology Provincial Group, (3) Post-COVID Condition Phys Med/Rehab Provincial Group, (4) Post-COVID Condition Chronic Fatigue Syndrome—Environ Health Provincial Group, and (5) Post-COVID Condition Respiratory Recovery Provincial Group.

Measures considered when recruiting specialists to eConsult BASE-managed groups include the specialist's experience, level of interest, and ability to maintain an adequate case volume, as well as maintaining equity across and within regions or organizations. The services leveraged existing eConsult specialists and recommendations from peers to identify eligible specialists.

In total, 19 specialists were available to respond to long COVID questions during the study period (5 on the Champlain BASE and 14 on the Ontario eConsult service) from various specialty backgrounds, including internal medicine, infectious disease, neurology, physiatry and respirology (Figure 1).

Figure 1. Timeline of the establishment of a long COVID specialty group for eConsult services. WHO: World Health Organization.



PCPs initiate an eConsult by logging onto a secure web portal and selecting the specific specialty service from a dropdown menu. The PCP fills out a structured electronic form and submits the case along with any pertinent attachments (eg, laboratory or imaging data). The case is assigned to a specialist from the chosen group, who responds within 7 days (the median specialist response time is typically 1 day [5]). When a case is closed, the PCP completes a mandatory close-out questionnaire that gathers information about the case's outcome, whether a referral was originally contemplated and whether it was ultimately avoided as a result of the specialist's advice, and the case's educational value and capacity to serve as material for continuing medical education by using a 5-point Likert scale (Multimedia Appendix 1).

Data Collection and Analysis

We computed descriptive statistics to examine the overall use of the long COVID specialist group from June 1, 2021, to July 31, 2022. Data from both the Ontario eConsult and Champlain eConsult BASE services were analyzed to determine the total number of cases submitted to the long COVID specialty group. Time stamps for each eConsult were assessed to examine the response interval for specialist responses, and the total self-reported time billed was used to determine the amount of time each specialist spent responding to the eConsult.

Following the established methodology, case transcripts of PCP questions from eConsults submitted through the Champlain eConsult BASE between June 1, 2021, and July 31, 2022, were retrospectively reviewed. Case-level data were accessible for Champlain eConsult BASE but not for Ontario eConsult. Two investigators (MQ and JS) reviewed cases retrospectively using 2 validated taxonomies: the International Classification for Primary Care, version 2 and Ely et al's taxonomy of generic clinical questions. The International Classification for Primary Care, version 2 was used to classify clinical topics and patient-presenting symptoms. The taxonomy of generic clinical question was used for question classification. In many cases, a single eConsult case had multiple questions, and patients often presented with multiple symptoms; therefore, each case was not restricted to a single clinical question or presenting complaint or symptom.

We performed a descriptive analysis of all eConsult close-out questionnaires completed by the PCPs. Specifically, we sought to determine the PCP's course of action after the eConsult (eg, face-to-face referral avoided and face-to-face referral still needed) and to gain insight into the nature of the advice given to the PCP (eg, new advice given and advice confirmed course of action initially contemplated).

Results

Overview

A total of 40 PCPs submitted 47 eConsults through Champlain eConsult BASE between June 1, 2021, and July 31, 2022, and 197 PCPs submitted 228 cases through Ontario eConsult between December 1, 2021, and July 31, 2022. These cases comprised 0.2% (n=23,067) of the total number of closed cases submitted to Champlain eConsult BASE and 0.5% (n=45,494)

of all Ontario eConsult closed cases during their respective time frames.

The median time for a specialist to respond to the eConsult was 0.6 (IQR 0.19 - 2.36; mean 1.7, SD 2.29) and the average time the specialist billed in responding to each eConsult was 27.5 (SD 18.53; median 20, IQR 15-45) minutes.

Clinical Content of eConsults

The retrospective taxonomy analysis assessed the clinical questions PCPs asked in the 47 cases submitted through Champlain eConsult BASE. Patients had a median age of 43 (IQR 6-65) and presented 24 unique symptoms across cases (Table 1).

Table . Percentage of patients with reported symptom or finding.

Symptom or finding	Patients (n=47), n (%)
Fatigue	14 (30)
Dyspnea	7 (15)
Altered smell (anosmia, parasomnia)	6 (13)
Cough	6 (13)
Cognitive changes	6 (13)
Chest pain	5 (11)
Palpitations	3 (6)
Altered taste (dysgeusia, ageusia)	3 (6)
Congestion	3 (6)
Alopecia	2 (4)
Myalgia	2 (4)
Rash	2 (4)
Paresthesia	2 (4)
Headache	2 (4)
Tremor	2 (4)
Insomnia	2 (4)
Fever	1 (2)
Eye pain	1 (2)
Tinnitus	1 (2)
Arthralgia	1 (2)
Back pain	1 (2)
Abdominal pain	1 (2)
Edema	1 (2)
Asthenia	1 (2)

The 5 most common symptoms were fatigue (14/47, 30%), dyspnea (7/47, 15%), cough (6/47, 13%), altered sense of smell (ie, anosmia and parosmia; 6/47, 13%), and cognitive changes (6/47, 13%). The five main question categories asked by PCPs were: (1) management of chronic symptoms of COVID-19, (2) need for additional work-up or follow-up testing, (3) community resources to support or manage patients with long COVID, (4) diagnostic clarification, and (5) guidance regarding COVID-19 vaccination (Table 2; sample questions in Multimedia Appendix 2). The most common topics pertained to guidance on the management of chronic symptoms of COVID-19 (22/47 cases, 47%), the need for additional work-up or follow-up testing (20/47 cases, 43%), suggestions for community resources to help support or manage patients with chronic COVID-19 (9/47 cases, 19%), and diagnostic clarification (7/47 cases, 15%).

Table . Content of clinical questions asked of post-COVlD-19 condition specialists by electronic consultation (eConsult).

Content topic	eConsult cases (n=47), n (%)
Management of chronic symptoms of COVID-19	22 (47)
Need for additional work-up or follow-up testing	20 (43)
Community resources to support or manage patients	9 (19)
Diagnostic clarification	7 (15)
Guidance related to COVID-19 vaccination	7 (15)
Necessary to refer to specialist	6 (13)
Interpretation of testing	4 (9)
Other	3 (6)

Outcome of eConsults

PCPs completed the post-eConsult questionnaire for 268 of the 275 cases submitted on both the Champlain eConsult BASE and Ontario eConsult services during the study timeframe. Through eConsult, PCPs were able to confirm a course of action that they had initially contemplated in 35% (95/268) of cases and received advice on a new or additional course of action in 59% (157/268) of cases (Table 3). In 38% (102/268) of cases, PCPs had initially contemplated a face-to-face referral but found it unnecessary after the eConsult, while an actual face-to-face referral was needed in only 24% (64/268) of cases (Table 4).

Table . Impact of electronic consultation (eConsult) on PCP[a] course of action.

PCP survey answer	Responses (n=268), n (%)
I got good advice for a new or additional course of action	157 (59)
I was able to confirm a course of action that I originally had in mind	95 (35)
I got good advice for a new or additional course of action that I am not able to implement[b]	5 (2)
I did not find the advice very useful[c]	5 (2)
Other	6 (2)

[a]PCP: primary care provider.

[b]Response option available only for the Champlain eConsult BASE Service.

[c]Response option available only for the Ontario eConsult Service.

Table . Impact of electronic consultation (eConsult) on need for face-to-face referral.

PCP[a] survey answer	Responses (n=268), n (%)
Referral was originally contemplated but now avoided at this stage	102 (38)
Referral was not originally contemplated and is still not needed	88 (33)
Referral was originally contemplated and is still needed	57 (21)
Referral was not originally contemplated, but eConsult process resulted in a referral being initiated	7 (3)
There was no particular benefit to using eConsult in this case[b]	2 (1)
Other	12 (4)

[a]PCP: primary care provider.

[b]Response option available only for the Ontario eConsult Service.

Discussion

Principal Results

This study demonstrates the considerable potential of an eConsult service to support PCPs navigating the complexity and diversity of long COVID symptoms. By facilitating quick, low-barrier access to specialist advice, eConsult empowered PCPs to provide the best possible care for patients with long COVID while avoiding long wait times for specialist referrals. PCPs were able to get prompt specialist advice from a specialist with a median response time of just over 12 hours. The main question types identified in the study related to (1) the management of chronic symptoms of COVID-19, (2) the need for additional work-up or follow-up testing, (3) community resources to support or manage patients with long COVID, (4)

diagnostic clarification, and (5) guidance regarding COVID-19 vaccination. A face-to-face referral was not needed after the majority of eConsult cases, so patients could be given an expert-guided treatment plan without waiting to see a specialist. To our knowledge, this is the first study to examine an innovative electronic solution to help support PCPs in managing long COVID.

Limitations

This study has several limitations. First, the taxonomy analysis used data from a single health region in Ontario, which may impact the generalizability of our findings. As mentioned above, we only had access to case-level data from the Champlain eConsult BASE service and not the Ontario eConsult service. The Champlain region is a culturally and linguistically diverse region and our findings were externally validated by specialist team members who answered eConsults from across Ontario and confirm the nature of the questions identified in this study are representative of common questions asked across the province. Furthermore, this study only examined eConsults submitted to the long COVID group and it is possible that relevant COVID-19 cases were submitted to other specialty groups such as infectious disease. Finally, neither service collects detailed demographic information about the PCPs using the service, so we were unable to report any of this information in our results section.

Comparison With Prior Work

With studies and data continuing to emerge about long COVID, there remain significant challenges in diagnosing and managing patients with this condition (O'Hare et al [9]). A systematic review conducted by Macpherson et al [10] found patients commonly expressed concerns about the lack of knowledge about long COVID, and often found they received conflicting or inconsistent advice from providers. There is currently no effective validated treatment for long COVID [11]. This challenge with management was evident in our findings, as approximately half of the questions PCPs posed were related to advice on management, and one-fifth inquired about community resources to support with management. Furthermore, 59% of PCPs surveyed reported receiving new advice that changed their management plans.

Recognizing the burden of long COVID, many jurisdictions worldwide have implemented multispecialty long COVID clinics and rehabilitation centers. Previous studies have shown that these rehabilitation programs can have a positive impact on various patient outcomes [12,13], such as increased quality of life, reduced fatigue, improved functional status, and better mental health [13]. Unfortunately, the demand for such clinics is high, with many reporting long wait times. Digital strategies such as eConsult provide a means for quicker access to support patients and initiate management. They also have the potential to reduce demand by allowing the PCP to provide guided care; approximately 40% of PCPs surveyed highlighted that a face-to-face referral was not needed as a result of the advice they received through eConsult. This may lessen the backlog of seeing long COVID specialists by decreasing the number of unnecessary in-person referrals. Additionally, we were able to quickly adapt the eConsult service to implement long COVID groups, whereas setting up specific specialty clinics requires greater coordination, time, and cost.

One of the other challenges in identifying and managing long COVID relates to its diversity of presentation; the Public Health Agency of Canada has reported over 100 different symptoms associated with long COVID. In the 47 eConsults examined in this study, 24 unique symptoms were reported, with the 5 most common being fatigue, dyspnea, cough, altered sense of smell (ie, anosmia and parosmia), and cognitive changes. These findings are in line with other international studies. For example, a UK-based study conducted by Carfi et al [14] also noted fatigue as the most common long COVID symptom (51%), followed by dyspnea (35%), arthralgia (25%), and concentration difficulties (25%). This diversity in presentations can lead to challenges with diagnosis, which is particularly important as the symptoms of long COVID overlap with other serious and possibly life-threatening complications associated with COVID-19, such as pulmonary embolism, myocarditis, and organizing pneumonia. Several publications have reported delays in the diagnosis of life-threatening conditions such as pulmonary embolism as they were mistaken for symptoms of COVID-19 (Yousefazai and Bhimaraj [15]; Melazzini et al [16]). In total, 15% (7/47) of PCP questions related to diagnostic clarification, while 43% (20/47) of eConsults asked about the need for additional testing to rule out other conditions. This highlights the importance of getting timely advice to confirm the diagnosis and avoid delays in the diagnosis of possibly life-threatening conditions. The diversity of presentations also demonstrates the need for a multidisciplinary approach to providing support for patients with long COVID. To address this need, Ontario eConsult established 5 unique subgroups that PCPs can select (ie, internal medicine, neurology, physical medicine or rehabilitation, chronic fatigue syndrome, and respirology) involving providers from various specialties to better support PCPs and their patients in getting helpful and appropriate advice in a timely manner.

Conclusions

The long COVID groups available through the Champlain eConsult BASE and Ontario eConsult services provided rapid access to a multispecialty service that facilitated the avoidance of unnecessary face-to-face referrals. Long COVID is a multisystemic condition that is often debilitating, with a significant impact on one's quality of life and mental health. Given the lack of knowledge around long COVID, limited timely access to specialized long COVID clinics, and the possibility of delayed diagnosis of life-threatening conditions associated with COVID-19 (eg, pulmonary embolism), there is an urgent need for innovative digital solutions such as eConsult to better support PCPs to provide patients with timely access to specialty advice. An assessment of eConsults highlighted 5 common question types, providing insight into potential gaps in knowledge among PCPs that could help guide medical education and policy.

Acknowledgments

This research received funding from Ontario Health.

Conflicts of Interest

CL and EK are the cofounders of the eConsultBASE service but have no commercial interest in the service. EK is the executive director of the Ontario eConsult Centre of Excellence and receives salary support from Ontario Health. EK completes occasional electronic consultations (eConsults) as a specialist through the service for which she is reimbursed. CL is the director of evaluation for the Ontario eConsult Centre and receives salary support from Ontario Health.

Multimedia Appendix 1
Mandatory close-out survey.
[DOCX File, 13 KB - humanfactors_v12i1e58582_app1.docx]

Multimedia Appendix 2
Sample questions asked of post-COVlD-19 condition specialists.
[DOCX File, 19 KB - humanfactors_v12i1e58582_app2.docx]

References

1. COVID-19 dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2025-02-20]
2. Coronavirus disease (COVID-19): post COVID-19 condition. World Health Organization. 2023. URL: https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-(covid-19)-post-covid-19-condition [accessed 2025-02-20]
3. Long-term symptoms in canadian adults who tested positive for COVID-19 or suspected an infection, january 2020 to august 2022. Statistics Canada. 2022. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/221017/dq221017b-eng.htm [accessed 2025-02-20]
4. Singh J, Garber GE, Keely E, Guglani S, Liddy C. Evaluation of an electronic consultation service for COVID-19 care. Ann Fam Med 2022;20(3):220-226. [doi: 10.1370/afm.2807] [Medline: 35606132]
5. Liddy C, Mitchell R, Guglani S, et al. The provincial spread and scale of the Ontario eConsult service: evaluation of the first 2 years. Ann Fam Med 2022;20(3):262-265. [doi: 10.1370/afm.2812] [Medline: 35606133]
6. Liddy C, Drosinis P, Armstrong CD, McKellips F, Afkham A, Keely E. Improving access to allied health professionals through the Champlain BASE. BMJ Open 2016;6(6):e010920. [doi: 10.1136/bmjopen-2015-010920]
7. Keely E, Williams R, Epstein G, Afkham A, Liddy C. Specialist perspectives on Ontario provincial electronic consultation services. Telemed e-Health 2019 Jan;25(1):3-10. [doi: 10.1089/tmj.2018.0012]
8. Liddy C, Sethuram C, Mihan A, Moroz I, Keely E. Primary care providers' perspectives on the Ontario eConsult program. Telemed J E Health 2021;27(9):1039-1045. [doi: 10.1089/tmj.2020.0338]
9. O'Hare AM, Vig EK, Iwashyna TJ, et al. Complexity and challenges of the clinical diagnosis and management of long COVID. JAMA Netw Open 2022 Nov 1;5(11):e2240332. [doi: 10.1001/jamanetworkopen.2022.40332] [Medline: 36326761]
10. Macpherson K, Cooper K, Harbour J, Mahal D, Miller C, Nairn M. Experiences of living with long COVID and of accessing healthcare services: a qualitative systematic review. BMJ Open 2022 Jan 11;12(1):e050979. [doi: 10.1136/bmjopen-2021-050979] [Medline: 35017239]
11. Davis HE, McCorkell L, Vogel JM, Topol EJ. Long COVID: major findings, mechanisms and recommendations. Nat Rev Microbiol 2023 Mar;21(3):133-146. [doi: 10.1038/s41579-022-00846-2] [Medline: 36639608]
12. Smith JL, Deighton K, Innes AQ, et al. Improved clinical outcomes in response to a 12-week blended digital and community-based long-COVID-19 rehabilitation programme. Front Med (Lausanne) 2023;10:1149922. [doi: 10.3389/fmed.2023.1149922] [Medline: 37293307]
13. Jimeno‐Almazán A, Franco‐López F, Buendía‐Romero Á, et al. Rehabilitation for post‐COVID‐19 condition through a supervised exercise intervention: a randomized controlled trial. Scand Med Sci Sports 2022 Dec;32(12):1791-1801. [doi: 10.1111/sms.14240] [Medline: 36111386]
14. Carfì A, Bernabei R, Landi F, Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent symptoms in patients after acute COVID-19. JAMA 2020 Aug 11;324(6):603-605. [doi: 10.1001/jama.2020.12603] [Medline: 32644129]
15. Yousefzai R, Bhimaraj A. Misdiagnosis in the COVID-19 Era: when zebras are everywhere, don't forget the horses. JACC Case Rep 2020 Aug;2(10):1614-1619. [doi: 10.1016/j.jaccas.2020.04.018] [Medline: 32342051]
16. Melazzini F, Reduzzi M, Quaglini S, Fumoso F, Lenti MV, Di Sabatino A. Diagnostic delay of pulmonary embolism in COVID-19 patients. Front Med (Lausanne) 2021;8:637375. [doi: 10.3389/fmed.2021.637375] [Medline: 33996852]

Abbreviations

eConsult: electronic consultation
PCP : primary care provider

Edited by A Kushniruk; submitted 21.03.24; peer-reviewed by K Peeters; revised version received 20.12.24; accepted 15.01.25; published 28.02.25.

Please cite as:
Singh J, Quon M, Goulet D, Keely E, Liddy C
The Utilization of Electronic Consultations (eConsults) to Address Emerging Questions Related to Long COVID-19 in Ontario, Canada: Mixed Methods Analysis
JMIR Hum Factors 2025;12:e58582
URL: https://humanfactors.jmir.org/2025/1/e58582
doi:10.2196/58582

Understanding Appropriation of Digital Self-Monitoring Tools in Mental Health Care: Qualitative Analysis

Lena de Thurah[1], MSc; Glenn Kiekens[1,2,3], PhD; Jeroen Weermeijer[1], PhD; Lotte Uyttebroek[1], MSc; Martien Wampers[1], PhD; Rafaël Bonnier[1], MSc; Inez Myin-Germeys[1], PhD

[1]Department of Neurosciences, Center for Contextual Psychiatry, KU Leuven, Herestraat 49 ON5B, bus 1029, Leuven, Belgium

[2]Research Unit of Clinical Psychology, Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium

[3]Department of Medical and Clinical Psychology, Tilburg University, Tilburg, Netherlands

Corresponding Author:
Lena de Thurah, MSc
Department of Neurosciences, Center for Contextual Psychiatry, KU Leuven, Herestraat 49 ON5B, bus 1029, Leuven, Belgium

Abstract

Background: Digital self-monitoring tools, such as the experience sampling method (ESM), enable individuals to collect detailed information about their mental health and daily life context and may help guide and support person-centered mental health care. However, similar to many digital interventions, the ESM struggles to move from research to clinical integration. To guide the implementation of self-monitoring tools in mental health care, it is important to understand why and how clinicians and clients adopted, adapted, and incorporated these tools in practice.

Objective: Therefore, this study examined how clinicians and clients within a psychiatric center appropriated an ESM-based self-monitoring tool within their therapy.

Methods: Twelve clinicians and 24 clients participated in the piloting of the ESM tool, IMPROVE. After utilizing the tool, 7 clinicians and 11 clients took part in semistructured interviews. A thematic framework analysis was performed focusing on participants' prior knowledge and expectations, actual use in practice, and potential future use of ESM tools.

Results: Many participants experienced that the ESM tool provided useful information about clients' mental health, especially when clinicians and clients engaged in collaborative data interpretation. However, clinicians experienced several mismatches between system usability and their technical competencies, and many clients found it difficult to comply with the self-assessments. Importantly, most participants wanted to use digital self-monitoring tools in the future.

Conclusions: Clinicians' and clients' choice to adopt and integrate self-monitoring tools in their practice seems to depend upon the perceived balance between the added benefits and the effort required to achieve them. Enhancing user support or redesigning ESM tools to reduce workload and data burden could help overcome implementation barriers. Future research should involve end users in the development of ESM self-monitoring tools for mental health care and further investigate the perspectives of nonadopters.

(JMIR Hum Factors 2025;12:e60096) doi:10.2196/60096

KEYWORDS

digital self-monitoring; technology appropriation; experience sampling method; mental health care; mental health; self-monitoring; digital health; adoption; implementation; thematic; usability; interview; experience; attitude; opinion; perception; perspective; acceptance

Introduction

To improve access to and facilitate person-centered mental health care, digital technologies are increasingly being deployed to collect and share health-related data, deliver care, and support individuals in managing their health [1,2]. The experience sampling method (ESM; also termed ecological momentary assessment) is a structured diary technique that enables individuals to collect detailed information about their mental health in their daily lives [3]. Using smartphone apps, the ESM prompts individuals to complete brief self-assessments multiple times daily, assessing their emotional states, behaviors, and psychological symptoms. By allowing individuals to collect this information and share it with their clinicians, the ESM can ensure that therapeutic decisions are aligned with clients' everyday experiences and needs. For example, by providing insights into which activities an individual commonly engages in, and how they respond to different daily life situations and stressors, clients and clinicians can make shared decisions about the focus of treatment [4].

Despite its potential, the ESM, similar to many other digital mental health interventions [5,6], struggles to move from

piloting and trialing to actual clinical implementation. To address this research-to-practice gap, scholars have made calls to adopt user-centered design strategies for developing digital mental health tools [7]. User-centered design emphasizes the importance of conducting in-depth analyses of users' goals, needs, and context of use to inform technology design [8]. Nonetheless, most evaluations of digital health technologies focus primarily on quantifiable, technical aspects such as system performance, usability scores, and cost benefits [9,10]. However, an often overlooked but crucial step in implementing the ESM and self-monitoring tools is understanding how and why people integrate these technologies into their daily practices [11]. For example, what are the goals people hope technologies will help them achieve, what difficulties do individuals encounter when using new technologies, and what strategies do they apply to overcome technology use obstacles?

Studying the situated use of the ESM can provide essential knowledge into how users appropriate and make sense of these tools [12,13]. Technology appropriation can be examined in three main stages [13]. The first stage involves users' prior knowledge and expectations, which determine their initial willingness to adopt the technology. Expectations about potential benefits (eg, simplifying tasks) and the effort required (eg, time spent on training) are particularly crucial in shaping the initial interest [14,15]. In the second phase, users explore and evaluate the technology's capabilities, while becoming familiar with its functions. They adapt their practices to integrate the technology (eg, maintaining an internet connection), and adapt the technology to suit their needs (eg, disabling certain features or altering settings) [13]. They will test the technology's applicability in different situations and determine what features are most useful to them. Finally, if users find that the technology has added value and helps them achieve their goals and tasks, they might develop new routines and workflows that allow them to integrate and use the technology in their daily practice, thus moving into the third stage of persistent use [13].

To date, the appropriation of ESM tools in health care has largely been overlooked. While one study has examined how individuals with back pain appropriated ESM tools for pain management [16], there is a lack of research in the field of mental health care. As a consequence, we have a limited understanding of how individuals with mental illnesses and their clinicians interact with and make sense of digital ESM tools, and what might lead them to adopt or abandon these technologies. However, understanding this is vital for facilitating successful implementation [15,17]. To address this knowledge gap, we undertook a pilot implementation study, using the ESM tool "IMPROVE," a clinical prototype tool informed by research examining clinicians' and clients' design preferences [18,19]. IMPROVE was used as a part of therapy following a 3-step intervention, in which clinicians and their clients could (1) personalize the tool to clients' specific problems and situations, (2) self-monitor clients' mental health and daily activities via an app, and (3) review the collected data in summarized graphs in an online dashboard. To inform the further development and implementation of clinical ESM tools, this paper aimed to evaluate participants' (1) prior knowledge and expectations, (2) actual use in practice, and (3) potential future integration of ESM-based self-monitoring tools. Addressing these gaps in the literature will provide valuable insights into why and how clients and clinicians in mental health care choose to adopt or abandon digital self-monitoring tools in therapy.

Methods

Ethical Considerations

All participants provided written informed consent, and all procedures were approved by the medical ethics committee of KU Leuven (S64244). Participants were given a study ID to ensure anonymity, and interviews were pseudonymized during the transcription by removing personally identifiable information and replacing them with pseudo codes. Finally, participants were allowed to freely use the IMPROVE tool after completion of the study but were not given additional compensation.

Study Design and Recruitment

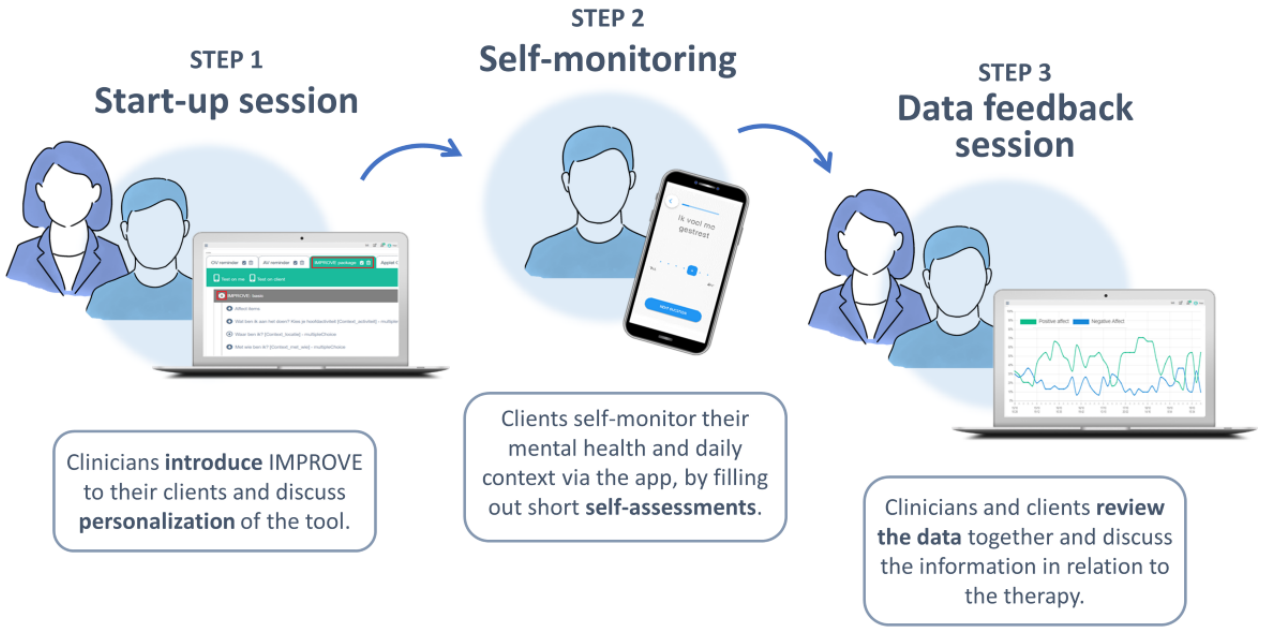
We conducted a pilot study within the University Psychiatric Center KU Leuven in Belgium in which we involved clients and clinicians in testing and evaluating IMPROVE. The research team was unable to access the clinic directly, due to COVID-19 restrictions. Therefore, invitations to participate were emailed to all psychiatrists and psychologists affiliated with the psychiatric center (N=142). The study aimed to enroll 12 clinicians, and initial invitations were followed by up to 2 reminder emails. For those interested, online informational sessions with the research team were organized. Participating clinicians were asked to recruit at least 1 client from their practice. By allowing clinicians to select clients, the referred sample provided a practical representation of the clients with whom clinicians were likely to use the tool within a real-life context.

The inclusion criteria were kept broad to reflect the diverse reality of clinical practice. To be eligible, clinicians needed to be certified mental health professionals and proficient in Dutch, which was the language the intervention was designed in. Clients had to be 18 years or older and also proficient in Dutch. To replicate real-world implementation, clients were asked to use their own smartphones. As a result, only clients owning a smartphone with at least 3G coverage were eligible to participate.

IMPROVE was run on the digital platform m-Path [20]. To enable clinicians to use IMPROVE, they received a manual with instructions on using the tool in their practice (Multimedia Appendix 1: IMPROVE training manual) and were invited to join an online training session. Clinicians and their clients were then requested to complete the 3-step intervention (Figure 1). In the first step, clinicians introduced IMPROVE to their clients and assisted them in downloading the app on their smartphones. They also discussed personalization of the tool, such as adding symptom-specific questions or adjusting the notification schedule. In the second step, clients self-monitored their mental health for 6 consecutive days. They received 10 semirandom notifications daily, prompting them to complete brief self-assessments on mood and context. Additionally, they received a morning notification to assess their sleep and an evening notification to evaluate the day. In the third step,

clinicians and clients reviewed the collected data together, exploring factors such as clients' activities, mood variability, and mood-context interactions.

Figure 1. Overview of the steps included in the IMPROVE intervention.



Data Collection and Analysis

Due to COVID-19 measures, data were collected remotely via online surveys in RedCap (Research Electronic Data Capture, Vanderbilt University) [21] and through Skype interviews (Version MSO; Skype Technologies). In some cases, poor connections affected the quality of the interview recordings; however, the overall data collection was not hindered by remote methods. Demographic information was collected at study enrollment and participants who completed the intervention were invited to participate in a semistructured interview (Multimedia Appendix 2: Interview guides). Interviews were audio recorded and transcribed verbatim. We performed a thematic analysis [22] in which we first conducted top-down content-coding of the transcripts using a thematic framework that identified interview material related to clinicians' and clients': (1) previous knowledge and expectations, (2) actual use in practice, and (3) potential future integration of the IMPROVE tool (Table 1). Hereafter, we performed additional inductive in vivo coding and used pattern and focused coding to create labels for the in vivo codes and compiled them into additional subthemes [23]. LdT undertook all primary and secondary coding, which was then revised by the coauthors and discussed in multiple peer debriefing sessions [24]. After each session, the coding categorizations and labels were modified and refined based on feedback from coauthors. To explore the potential impact of the different themes on the discontinued use of IMPROVE, additional information was extracted from informal participant contact records kept by the research team.

Table . Thematic framework for top-down coding of interviews.

Theme	Description
Prior knowledge and expectations	Participants talk about any prior experiences or knowledge they have regarding the ESM[a] or self-monitoring tools (analog or digital). Participants talk about their motivations and expectations related to using the ESM or self-monitoring tools and participating in the study.
Actual use in practice	Participants talk about how they used the IMPROVE tool and how they judged these experiences. Mentioning eg, what purpose they used it for, when, and how often they used it.
Potential future integration	Participants talk about their interest in using the IMPROVE tool (or similar tools) in the future. Participants talk about potential changes that they find would be relevant for future use and integration of the IMPROVE tool.

[a]ESM: experience sampling method.

Results

Participant Characteristics

Nineteen clinicians expressed interest in participating in the study, of whom 12 were enrolled. While a systematic record of the number of clients approached by clinicians was not maintained, at least 29 clients were invited, with 24 agreeing to participate. Of the enrolled participants, 8/12 clinicians (67%) and 17/24 clients (71%) completed the intervention; among these, 7/8 clinicians (88%) and 11/17 clients (65%) subsequently participated in an interview. A demographic summary of interview participants can be consulted in Multimedia Appendix 3: Demographic summary.

Thematic Analysis

A complete overview of the themes and subthemes of the analysis and how frequently they were mentioned by participants can be found in Multimedia Appendix 4: Overview of themes. Tabulated in vivo code summaries of participants' individual experiences can be found in Multimedia Appendix 5: Individually summarized experiences. Additionally, illustrative summaries are displayed in text boxes.

Previous Experience and Expectations

Individuals' experiences and expectations are important indicators of their initial willingness to adopt technologies. While only 1 clinician had experience using digital self-monitoring tools, most clinicians and some clients had experience using analog self-monitoring or diary techniques. Some clinicians also had experience with different forms of digital tools within their practice (eg, online training platforms, video consultations, and virtual reality). Similarly, several clients had experience using mental health apps, eg for practicing breathing techniques and cognitive-behavioral therapy exercises. Despite their limited experience, clinicians generally expected that IMPROVE would offer benefits over analog registration methods, allowing them to more easily collect and summarize information about their clients. However, several clinicians also anticipated that there would be challenges and limitations for using the tool. The most commonly mentioned was that low digital literacy could be a barrier for both clients and clinicians (Clinician 0100 Textbox 1). Clinicians' motivation for testing the tool was primarily driven by curiosity and the conviction that they need to master digital tools, as these are becoming increasingly dominant in the health care sector. Interestingly, most clients were primarily motivated by a wish to help and contribute to research but also expressed a curiosity toward what digital mental health tools might offer (Client 0501 Textbox 2). Some clients hoped that IMPROVE would allow them and their clinicians to get a better understanding of their mental health problems.

Textbox 1. "Might use digital self-monitoring tools in the future"– selected participants' experiences.

CLINICIAN 0100:

Prior knowledge and expectations

- I had no experience with digital self-monitoring, but I am currently testing VR technology in therapy. I have previously worked with analog self-registration in therapy.
- I expected it would be technically challenging to work with the tool.
- I want to try using new methods in therapy and I was eager to try the tool.

Actual use in practice

- It was difficult to find things in the dashboard and I needed help from the research team to use the tool. After a while using the platform got easier, but it required a lot of time to get started.
- It was difficult to draw conclusions about my clients' data, but I asked my client for clarifications when I was not able to interpret their data.

Potential future integration

- The tool gave me more information about what happens in clients' lives.
- I might use the tool again if it is made easier to use.

CLIENT 0503:

Prior knowledge and expectations

- I had no experience with self-monitoring apps.
- I hoped to gain more insight into my mental health.

Actual use in practice

- Assessment frequency was high, but okay for 1 week.
- I think I responded to almost all notifications, but I sometimes missed notifications because I forgot my phone. Also, when I was not feeling well I didn't respond to the notifications.
- Identifying, labeling, and scoring emotions on a scale are difficult.
- My therapist had difficulties operating the dashboard, so we didn't go into detail with the data.

Potential future integration

- I might use the tool again, but it is tiring to do for a long time.
- I prefer using the tool as a part of therapy, but I would like to have access to the data myself.

CLIENT 0802:

Prior knowledge and expectations

- I have no experience with health apps, but I use analog methods to keep track of my mental health.
- Digital tools allow you to do more than analog tools, so I thought it would be interesting to try.

Actual use in practice

- The number of notifications was okay, but I sometimes missed notifications because I was working or sleeping.
- Using the tool during the therapy session made it easier to recollect things that had happened.

Potential future integration

- I might consider using the tool again.

Textbox 2. "Not likely to use digital self-monitoring tools in the future"– selected participants' experiences.

CLIENT 0501:

Prior knowledge and expectations

- I had no experience with self-monitoring apps.
- I was hesitant to participate because I do not like self-assessments.
- I thought it was worth trying and I wanted to contribute to improving mental health care.

Actual use in practice

- It was difficult always to remember to have my phone with me.
- Notifications were sometimes disturbing and the assessment frequency was too high.
- It was important to me that I did not miss notifications; therefore, fear of missing notifications would sometimes stress me.
- Self-reflection was sometimes difficult and confronting.
- It was interesting to look at my data.

Potential future integration

- I don't find it likely that I will use the tool again.

Actual Use in Practice

Participants' willingness to test the IMPROVE tool was influenced by their initial expectations that digital self-monitoring would provide them with benefits and advantages. Below we describe how clinicians and clients used the tool in practice and how they evaluated its capabilities and contextual fit.

Step 1: Start-Up Session

Clinicians generally found that the time investment required to learn to use IMPROVE was too much considering the time they had available. Experiencing time constraints was also one of the most common reasons reported by clinicians who did not complete the intervention. Personalizing clients' questionnaires in particular asked for extra time and effort from clinicians. Despite emphasizing that the personalization of the IMPROVE questionnaires made the tool more relevant, clinicians made limited use of the available personalization options. Most clients also expressed that personalization of the tool is desirable but that clinicians did not discuss this with them.

Furthermore, while most clinicians were convinced that using the IMPROVE tool would become easier with practice, many clinicians indicated that the complexity of the tool was too high. In particular, clinicians found that navigating the dashboard was not intuitive and that it was difficult to find things and set up the clients' questionnaires. Thus, many experienced that they did not have adequate competencies to make full use of the tool; "I constantly felt I was doing something wrong," 1 clinician explained. However, at the same time, clinicians made limited use of the training manual and expressed a wish for more in-person support. This was also reflected by several clinicians contacting the research team for additional support after the initial training (Clinician 0100,Textbox 1). Interestingly, some clinicians resolved their need for support by organizing training sessions with colleagues.

Step 2: Self-Monitoring

The majority of clients reported that they made an effort to comply with the notifications, but many found that the frequency of the self-assessments (10 notifications per day) was too high. A few clients indicated that they would consciously skip assessments because they did not consider it important to respond to all notifications. Conversely, some clients also mentioned that missing notifications would make them feel guilty and annoyed, or that the thought of missing assessments made them nervous and stressed (Client 0501, Textbox 2). Although clients reported that completing the assessments did not take them long, many experienced difficulties responding to the notifications within the fixed 15-minute response window. Several reasons for this were voiced, of which being occupied with work and other daily activities was the most frequent (Client 0802, Textbox 1; Client 0601, Textbox 3). Having an irregular day rhythm or sleep schedule that did not match the notification schedule, needing to carry one's phone and ensure an internet connection, and feeling unwell or tired were also factors that made complying difficult. Importantly, clients who dropped out of the study also reported a lack of time, difficulties complying with the assessments, notifications stress, and feeling unwell as reasons for not completing the intervention.

Textbox 3. "Likely to use digital self-monitoring tools in the future"– selected participants' experiences.

CLINICIAN 0600:

Prior knowledge and expectations

- I have no experience with digital self-monitoring, but I have used analog self-monitoring techniques in therapy.
- I think digital tools are easier to use than analog self-monitoring methods. Digital tools are used more and more, and we need to offer clients tools that can help them achieve their goals.

Actual use in practice

- It takes time and effort to learn to use the tool.
- I didn't personalize my client questionnaire, because I wanted to start with the basics.
- I monitored my clients' responses and contacted them if I could see they weren't responding.
- There was a lot of data; I started with simpler visualizations and gradually added things. If I didn't know the client beforehand, I would not be able to make sense of the data.

Potential future integration

- The tool provided more details and an overview and allowed us to go more in-depth with problems that we already knew were there. We used the tool to identify what was important for the client.
- I will certainly consider using this tool again with my clients.

CLIENT 0601:

Prior knowledge and expectations

- I had no experience with health apps.
- I had no specific expectations and I'm generally skeptical about health apps, but I wanted to help research.

Actual use in practice

- I responded to as many notifications as I could, but I missed a lot because I was busy. I didn't consider it a problem that I missed some notifications. I also had to remember to bring my phone and connect it to 4G.
- It is sometimes difficult to assess whether your mood changed and how much.
- I was amazed to see what came out of the data. My therapist was able to do a lot with the data, despite my low compliance.

Potential future integration

- The tool provided useful information and allowed me to get to know myself better.
- I would be interested in using the tool again, but I would like to have access to the data myself.

CLINICIAN 1200:

Prior knowledge and expectations

- I used digital self-monitoring tools before in therapy.
- Digital self-monitoring tools are easier to use and give better insight into clients' lives than other methods. However, self-monitoring can be difficult for some clients.

Actual use in practice

- Setting up the tool took a lot of time.
- It was interesting to look at the data, but the many visualizations were somewhat overwhelming.

Potential future integration

- The tool quickly provides you with an overview of how your client is doing.
- I would like to use the tool again, but it should be made less burdensome for clients. It would be interesting to monitor clients for a longer period to evaluate their progress and the effect of treatments.

CLIENT 1201:

Prior knowledge and expectations

- I had no experience with mental health apps.

- I expected the tool would make it easier for my therapist to understand my problems.
- I think there is a need for more mHealth in mental health care and that this tool might help me and others.

Actual use in practice

- There are too many notifications and they are sometimes disturbing. People should be allowed to snooze notifications if they are busy.
- I did my best to respond to the notifications. On good days, it is okay to complete the assessments, but on bad days, it is difficult.
- My therapist identified moments when I was feeling bad and tried to understand these.

Potential future integration

- The tool can make you aware of what you need to work on, and it makes it easier to monitor the effects of your treatment.
- I would like to keep using the tool to check how I am doing.

Several clients found the practice of labeling and rating their emotions difficult. Some indicated that they were unsure whether they completed the self-assessments correctly and expressed a need for more guidance from their clinicians on how to respond to the assessments. Several clients also expressed a wish for more open questions that would allow them to describe their experiences in more detail, as they did not find that the default questions were sufficiently able to capture their experiences. Interestingly, while clinicians were not instructed to do so, some monitored their clients' responses during the self-monitoring week as a form of remote monitoring of clients (Clinician 0600, Textbox 3).

Step 3: Data Feedback Session

To get the most out of the data feedback session, most clinicians reviewed their clients' data in preparation for the session. Many felt that preparation was necessary for them to understand the data and be able to discuss it with clients. Many clinicians found it challenging to interpret and navigate the graphs, which they often perceived as overwhelming. This generally led clinicians to be selective in the data they would review and discuss with clients. All clinicians reviewed the data with their clients and some also encouraged their clients to give their interpretation of the data or to provide further clarification (Clinician 0100, Textbox 1*)*. Similarly, clients indicated that the support of a clinician was crucial for them to make sense of their data. While most clients reported that they could understand the graphs, they also indicated relying on their clinician to help them interpret the data. These findings indicate that, despite challenges, most clinicians could provide useful feedback to their clients based on the self-monitoring data.

Potential Future Integration

Despite encountering challenges, both clinicians and clients described that the IMPROVE tool provided them with useful insights. Next, we will explore how these experiences influenced participants' willingness to integrate and use digital self-monitoring tools in their future practice. Clinicians and clients were generally open to using digital self-monitoring tools again. Participants who expressed the greatest interest in continued use more often experienced an added value of using IMPROVE (see Textbox 3). This included having access to more detailed information about clients' mental health, which helped create an overview and clarity, and identify focus points to discuss in therapy. Clinicians who had initial positive expectations and believed digital tools were important in supporting mental health care were also more willing to use digital self-monitoring again. The clinicians who were more hesitant reported more difficulties using IMPROVE and expressed a greater need for support (see Textbox 1).

Clients who expressed a low interest in using digital self-monitoring tools in the future also reported negative reactivity in the form of stress and increased negative emotions during self-monitoring (see Textbox 2). Interestingly, all these clients had high compliance during the self-monitoring. However, clients overall reported that the assessment frequency should be lowered to make future self-monitoring less burdensome. Moreover, some clients and clinicians expressed an interest in using the tool for extended periods to monitor progress and the effects of therapy. Finally, clients and clinicians also expressed a desire to adapt the self-assessment further and monitor factors more specific to the individual client.

Discussion

Principal Findings and Implications

To guide the successful implementation of digital self-monitoring tools in mental health care, it is pivotal to understand how people react, respond, and adapt to these technologies. We examined how clinicians and clients in a psychiatric center appropriated the ESM-based self-monitoring tool IMPROVE within their therapeutic practices. In line with existing theories, our results indicate that clinicians' and clients' willingness to adopt and integrate digital self-monitoring tools into their practice depend on the perceived balance between added benefits and the effort necessary to accomplish these [17]. The main benefits identified in our study included better information about clients' mental health, while system usability and assessment burden demanded extra effort from participants. Below we discuss the potential implications of this for the implementation of the ESM tools in clinical practice.

While some struggled more than others, all interviewed participants managed to reach a basic level of use of the IMPROVE tool within the intervention period. Furthermore, we observed a great interest and curiosity towards digital mental health tools, especially among clinicians. Many clinicians experienced that using IMPROVE added value and generated more clarity and focus in the therapy, by providing more detailed

information about the clients' mental health. Several clients also experienced these benefits. Our findings suggest that a collaborative effort between clients and clinicians is key to maximizing the experienced benefits of the ESM tools. While many clinicians struggled to construct a frame for the interpretation of clients' data, several turned to their clients to help make sense of the data. Clients also stressed the importance of having a clinician support them in interpreting their data and doubted whether they would be able to do this themselves. Other studies similarly highlighted the value of patients providing clinicians with additional information about patient-generated health data to unveil the subjective meanings of the data [25]. This indicates that collaborative data interpretation is an essential component in the clinical application of the ESM tools, which boosts the perceived usefulness of the tool. Therefore this component should be emphasized and strengthened in future implementation initiatives.

Most participants in our study expressed willingness to use the ESM or similar tools in the future, although many emphasized the necessity of making changes and improvements to the tool. This aligns with previous findings that clinicians and clients are generally interested in the ESM-based self-monitoring [18,19,26]. However, the need for further adaption was reflected by participants encountering several challenges in using and integrating IMPROVE into their practice. Many clinicians experienced difficulties navigating the tool's functionalities, which made it difficult for them to start using the tool and apply more advanced features, such as personalizing clients' questionnaires. Clinicians, therefore, expressed a need for more in-person support. Unfortunately, these types of usability issues are common in digital health tools [27]. A study examining usability problems of eHealth applications found that system navigation, interface design, and lack of built-in guidance and support accounted for 69% of users' usability issues [27]. Furthermore, consistent with other research [28,29], our study indicated that technology literacy and attitude can influence willingness to use digital tools. More anthropological field research on mental health care workers' existing technology habits and literacy might help inform the system design of the ESM tools and understand users' context and capabilities. Similarly, engaging end-users in co-design processes could help tackle usability issues that could lead users to abandon the tools. Finally, our results show that adequate support for clinicians and clients is needed to facilitate the implementation of the ESM tools in mental health care. Especially in the early adoption phases, users should be supported in familiarizing themselves with the tools and integrating them into their work routines. Potential solutions could include more built-in guidance functions in the tool or establishing additional structures (eg, service centers) that can provide direct user support.

Furthermore, our findings emphasize that self-monitoring demands a lot of clients, and can be difficult and burdensome for people with mental health problems. While clients expressed a wish to comply with the self-assessments, they often found it practically challenging. In line with other self-monitoring studies [30], we found that competing activities (eg, working) and technical issues (eg, device or internet access) were the most common reasons for missing assessments. To comply with the assessments, several clients had to change their phone habits, eg, ensuring that they always had their phone with them, and had notifications and internet connection on. As it is known that existing habits are important predictors of technology acceptance [17], this need to change habits can become a threat for sustained clinical implementation of ESM tools. Furthermore, some clients reported negative reactivity to self-monitoring, which was associated with less interest in using the tool again. Other studies also found that while self-monitoring can be motivating and helpful, it can also become a stressful activity that clients feel obliged to comply with [31]. This highlights the need to investigate how self-monitoring tools can be made less burdensome for users while still producing valuable information. Potential solutions could include the integration of passive monitoring [32,33] or adaptive assessment schemes that allow for periods with lower and higher assessment intensity [34,35].

Limitations

The findings of the study should be interpreted considering the following limitations. One limitation is that the study relied on referral sampling, which might have influenced the representativeness and diversity of the sample. Furthermore, the study was conducted in a large psychiatric center of a university in Belgium. Clients and clinicians within different health care settings or with different demographic backgrounds might have different experiences of using digital self-monitoring tools. Another limitation is the potential overrepresentation of people with a positive attitude towards digital mental health tools. Several clinicians for example recruited clients who they expected were interested in using digital tools and had the necessary technical literacy. Furthermore, none of the participants who dropped out of the intervention participated in the interviews. However, using contact records data, we established that some of the challenges experienced by participants who were interviewed were also reported as reasons for dropout by participants who did not complete the intervention. Future research should aim to better understand the views of nonadoptors, identify the main reasons for nonadoption, and work towards tackling these barriers in the design and implementation of ESM self-monitoring tools.

Conclusion

ESM-based self-monitoring can provide clinicians and clients with useful information about how clients' daily life activities correspond to fluctuations in their mental health. These benefits seem to increase when clinicians and clients engage in collaborative interpretation and sense-making of the clients' data. Therefore, ESM tools have a clear potential to support a person-centered approach to mental health care rooted in clients' daily life experiences. However, our study highlights that the effort required by clinicians and clients to integrate ESM tools into daily practice remains substantial due to system usability challenges and the burden of repeated assessments. Addressing these issues in future ESM tool developments will be crucial for successful implementation. One potential solution to tackle these barriers is enhancing user support. Another is modifying key features of ESM tools to better suit users and their contexts, such as reducing the frequency of assessment and data information load. Engaging end-users directly in this process

through user-centered design approaches may ensure a better fit between the tools, their context of use, and user goals and ultimately facilitate better adoption.

Acknowledgments

The authors would like to thank all clinicians and clients for participating in the study and contributing to this work. We also thank Merijn Mestdagh and Stijn Verdonck for programming the necessary software functionalities on the m-Path software platform. In addition, the authors would like to thank Silke Apers, Daphne Tuyaerts, and Tessa Biesemans for their contribution to data collection and Ana Teixeira and Daphne Tuyaerts for their help in designing the study. Finally, the authors would like to thank the Leuven IMMERSE patient advisory board members for providing feedback on the interpretation of the study results. This project received funding from the European Union's Horizon 2020 research and innovation program (777084; DynaMORE project and 945263; IMMERSE). GK is supported by the Research Foundation–Flanders (12ZZM21N and 1204924N).

Data Availability

To avoid potential identification of study participants, the complete interview transcripts will not be made publicly available.

Authors' Contributions

LT, GK, JW, MW, and IMG were involved in designing the study and setting up the data collection. Data analysis was undertaken in a collaboration between LT and LU. The first manuscript draft was written by LT with the support of GK. All authors contributed to the critical review of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
IMPROVE training manual.
[DOC File, 4491 KB - humanfactors_v12i1e60096_app1.doc]

Multimedia Appendix 2
Interview guides.
[DOC File, 35 KB - humanfactors_v12i1e60096_app2.doc]

Multimedia Appendix 3
Demographic summary.
[DOC File, 25 KB - humanfactors_v12i1e60096_app3.doc]

Multimedia Appendix 4
Overview of themes.
[DOC File, 30 KB - humanfactors_v12i1e60096_app4.doc]

Multimedia Appendix 5
Individually summarized experiences.
[DOC File, 40 KB - humanfactors_v12i1e60096_app5.doc]

References

1. Lupton D. The digitally engaged patient: Self-monitoring and self-care in the digital health era. Soc Theory Health 2013 Aug;11(3):256-270. [doi: 10.1057/sth.2013.10]
2. World Health Organization. World Mental Health Report: Transforming Mental Health for All 2022.
3. Myin-Germeys I, Kasanova Z, Vaessen T, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018 Jun;17(2):123-132. [doi: 10.1002/wps.20513] [Medline: 29856567]
4. van Os J, Verhagen S, Marsman A, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
5. Peiris D, Miranda JJ, Mohr DC. Going beyond killer apps: building a better mHealth evidence base. BMJ Glob Health 2018;3(1):e000676. [doi: 10.1136/bmjgh-2017-000676] [Medline: 29527353]

6. Torous J, Myrick K, Aguilera A. The need for a new generation of digital mental health tools to support more accessible, effective and equitable care. World Psychiatry 2023 Feb;22(1):1-2. [doi: 10.1002/wps.21058] [Medline: 36640397]
7. Mohr DC, Weingardt KR, Reddy M, Schueller SM. Three problems with current digital mental health research... and three things we can do about them. Psychiatr Serv 2017 May 1;68(5):427-429. [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
8. Gulliksen J, Göransson B, Boivie I, Persson J, Blomkvist S, Cajander Å. Key principles for user-centred systems design. In: Seffah A, Gulliksen J, Desmarais MC, editors. Human-Centered Software Engineering — Integrating Usability in the Software Development Lifecycle: Springer Netherlands; 2005:17-36. [doi: 10.1007/1-4020-4113-6_2]
9. Weermeijer JDM, Wampers M, de Thurah L, et al. Usability of the experience sampling method in specialized mental health care: pilot evaluation study. JMIR Form Res 2023 Nov 21;7:e48821. [doi: 10.2196/48821] [Medline: 37988137]
10. Simons CJP, Drukker M, Evers S, et al. Economic evaluation of an experience sampling method intervention in depression compared with treatment as usual using data from a randomized controlled trial. BMC Psychiatry 2017 Dec 29;17(1):415. [doi: 10.1186/s12888-017-1577-7] [Medline: 29284448]
11. Fidock J, Carroll J, Rynne A. Evaluating information systems: an appropriation perspective. In: Hart DN, Gregor SD, editors. Information Systems Foundations: The Role of Design Science: ANU Press; 2010:121-142 URL: http://www.jstor.org/stable/j.ctt24h3z6.9 [doi: 10.22459/ISF.12.2010.06]
12. Bal M, Benders J, Vermeerbergen L. "Bringing the covert into the open": a case study on technology appropriation and continuous improvement. Int J Environ Res Public Health 2022 May 23;19(10):6333. [doi: 10.3390/ijerph19106333] [Medline: 35627869]
13. Carroll J. Completing design in use: closing the appropriation cycle [abstract]. Presented at: 12th ECIS; 2004; Turku, Finland URL: http://aisel.aisnet.org/ecis2004/44 [accessed 2025-02-18]
14. Chiu TML, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Med Inform Decis Mak 2010 Nov 23;10:73. [doi: 10.1186/1472-6947-10-73] [Medline: 21092275]
15. Fidock JJ, Carroll J. The Model of Technology Appropriation: A Lens for Understanding Systems Integration in A Defence Context: ACIS; 2006.
16. Kaur E, Delir Haghighi P, Cicuttini FM, Urquhart DM. Smartphone-based ecological momentary assessment for collecting pain and function data for those with low back pain. Sensors (Basel) 2022 Sep 19;22(18):7095. [doi: 10.3390/s22187095] [Medline: 36146442]
17. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
18. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of patients and clinicians on the promise of the experience sampling method for psychiatric care. Psychiatr Serv 2019 Nov 1;70(11):983-991. [doi: 10.1176/appi.ps.201900050] [Medline: 31434558]
19. Weermeijer J, Kiekens G, Wampers M, Kuppens P, Myin-Germeys I. Practitioner perspectives on the use of the experience sampling software in counseling and clinical psychology. Behav Inf Technol 2024 Feb 17;43(3):540-550. [doi: 10.1080/0144929X.2023.2178235]
20. Mestdagh M, Verdonck S, Piot M, et al. m-Path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. Front Digit Health 2023;5:1182175. [doi: 10.3389/fdgth.2023.1182175] [Medline: 37920867]
21. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
22. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
23. Saldana J. The Coding Manual for Qualitative Researchers: SAGE Publications; 2012.
24. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017;16(1):1609406917733847. [doi: 10.1177/1609406917733847]
25. West P, Van Kleek M, Giordano R, Weal MJ, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. 2018 Apr 21 Presented at: CHI '18; Apr 21-26, 2018; Montreal QC Canada p. 1-13 URL: https://dl.acm.org/doi/proceedings/10.1145/3173574 [doi: 10.1145/3173574.3174058]
26. de Thurah L, Kiekens G, Sips R, Teixeira A, Kasanova Z, Myin-Germeys I. Using experience sampling methods to support clinical management of psychosis: the perspective of people with lived experience. Psychiatry Res 2023 Jun;324:115207. [doi: 10.1016/j.psychres.2023.115207] [Medline: 37087818]
27. Broekhuis M, van Velsen L, Peute L, Halim M, Hermens H. Conceptualizing usability for the eHealth context: content analysis of usability problems of eHealth applications. JMIR Form Res 2021 Jul 27;5(7):e18198. [doi: 10.2196/18198] [Medline: 34313594]
28. Kabakus AK, Bahcekapili E, Ayaz A. The effect of digital literacy on technology acceptance: an evaluation on administrative staff in higher education. Journal of Information Science 2023. [doi: 10.1177/01655515231160028]
29. Jang M, Aavakare M, Nikou S, Kim S. The impact of literacy on intention to use digital technology for learning: a comparative study of Korea and Finland. Telecomm Policy 2021 Aug;45(7):102154. [doi: 10.1016/j.telpol.2021.102154]

30. Ziesemer K, König LM, Boushey CJ, et al. Occurrence of and reasons for "missing events" in mobile dietary assessments: results from three event-based ecological momentary assessment studies. JMIR Mhealth Uhealth 2020 Oct 14;8(10):e15430. [doi: 10.2196/15430] [Medline: 33052123]
31. Wienroth M, Lund Holm Thomsen L, Høstgaard AM. Health technology identities and self. Patients' appropriation of an assistive device for self-management of chronic illness. Sociol Health Illn 2020 Jun;42(5):1077-1094. [doi: 10.1111/1467-9566.13079] [Medline: 32157709]
32. Medich M, Cannedy SL, Hoffmann LC, et al. Clinician and patient perspectives on the use of passive mobile monitoring and self-tracking for patients with serious mental illness: user-centered approach. JMIR Hum Factors 2023 Oct 24;10:e46909. [doi: 10.2196/46909] [Medline: 37874639]
33. Saunders KEA, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar;41:115-121. [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
34. Bucci S, Gumley A, Walters J, et al. CONNECT: using electronic devices (e.g. smartphones, smartwatches) to predict relapse of psychosis - research protocol. ISRCTN Registry 2024. [doi: 10.1186/ISRCTN24746936]
35. King ZD, Yu H, Vaessen T, Myin-Germeys I, Sano A. Investigating receptivity and affect using machine learning: ecological momentary assessment and wearable sensing study. JMIR Mhealth Uhealth 2024 Feb 7;12:e46347. [doi: 10.2196/46347] [Medline: 38324358]

Abbreviations

ESM: experience sampling method

Edited by A Kushniruk; submitted 03.05.24; peer-reviewed by K Mouloudj, M Galani; revised version received 06.12.24; accepted 23.12.24; published 03.03.25.

Please cite as:
de Thurah L, Kiekens G, Weermeijer J, Uyttebroek L, Wampers M, Bonnier R, Myin-Germeys I
Understanding Appropriation of Digital Self-Monitoring Tools in Mental Health Care: Qualitative Analysis
JMIR Hum Factors 2025;12:e60096
URL: https://humanfactors.jmir.org/2025/1/e60096
doi:10.2196/60096

Development and Testing of the Kids Hurt App, a Web-Based, Pain Self-Report App for First Nations Youths: Mixed Methods Study

Karlee Francis[1,2*], BSc; Julie Francis[3], RN; Margot Latimer[4,5*], RN, PhD; Hayley Gould[6], BA; Shante Blackmore[2,7*], BSc, MD; Emily MacLeod[5*], MA

[1]Eskasoni Health Centre, Eskasoni, Nova Scotia, Canada
[2]Dalhousie University, Halifax, Nova Scotia, Canada
[3]Tajikeim k Mi'kmaw Health & Wellness, Membertou, Nova Scotia, Canada
[4]School of Nursing, Dalhousie University, 6299 South Street, Halifax, Nova Scotia, Canada
[5]IWK Health, Halifax, Nova Scotia, Canada
[6]Nova Scotia Health, Sydney, Nova Scotia, Canada
[7]Nova Scotia Health, Halifax, Nova Scotia, Canada
[*]these authors contributed equally

Corresponding Author:
Margot Latimer, RN, PhD
School of Nursing, Dalhousie University, 6299 South Street, Halifax, Nova Scotia, Canada

Abstract

Background: First Nations children and youths may have unique ways to convey their health needs that have not been recognized by health providers. This may contribute to the disparity between high rates of mental health and physical pain and low rates of treatment for the conditions they experience. Evidence suggests that a colonial history has resulted in poor experiences with the health care system, lack of trust with health providers, and miscommunication between clinicians and patients. Contemporary ways, using both Indigenous and Western knowledge, are needed to bridge the gap in communicating pain.

Objective: The aim of this qualitative study was to test the usability and clinical feasibility of the Kids Hurt App with First Nations youths and clinicians working with youths.

Methods: Using a Two-Eyed Seeing approach, the Kids Hurt App was developed using concepts from validated mood and pain assessment apps combined with community-based research that gathered First Nations youths and clinicians perspectives on quality, intensity, and location of pain and hurt. The Kids Hurt App contains 16 screens accessible on any web-based device.

Results: In total, 3 rounds of low-fidelity testing (n=19), 2 rounds of high-fidelity testing (n=20), and 2 rounds of clinical feasibility testing (n=10) were conducted with First Nations youths (10 - 19 years) to determine the relevance, validity, and usability of the Kids Hurt App. High-fidelity testing was also conducted with 15 clinicians after completing the high-fidelity youth sessions. Youths had constructive suggestions that were used to improve the app in subsequent rounds of version testing. There was one main discrepancy between youths and clinicians related to preference for how best to visually convey pain. The youth's preference was maintained in the app.

Conclusions: All youths in all rounds of testing indicated that they would use the Kids Hurt App if it was available to them in a health care setting, with most clinicians noting that the app would be useful in practice.

(JMIR Hum Factors 2025;12:e48370) doi:10.2196/48370

KEYWORDS

app; eHealth; pain; Indigenous; First Nations; children; youths; mobile phone

Introduction

Overview

Indigenous peoples in Canada are comprised of 3 distinct groups: First Nations, Inuit, and Métis. They reside across Canada, from large cities to small isolated communities [1]. In comparison to non-Indigenous peoples, Indigenous peoples are the fastest-growing cohort in Canada with a significantly younger population [1]. Indigenous children younger than 14 years of age represent 25.4% of the total Indigenous population, while the same age group in the non-Indigenous population only compromises 16% [1]. Repeated reports of racism in health care, such as those in the In Plain Sight report describing Joyce

Echaquan's fatal health care experience in Quebec [2], combined with research findings [3,4] suggest that health care providers need to understand the structural and historical factors that contribute to the current racial disparities faced by Indigenous peoples [5]. A legacy of colonialism and present-day racism is a determinant of health that is unique to Indigenous peoples so efforts need to be made to create better understanding, appropriate assessment, and improved care for health conditions such as pain [6].

While 1 in every 5 Canadian people experiences chronic pain, this number is significantly higher in Indigenous communities, and most types of pain are more prevalent in Indigenous peoples [7]. Indigenous children are particularly vulnerable to experiencing pain. In one study examining pain diagnoses, a large sample of First Nations children (n=2631) who were compared to an age-, sex-, and location-matched sample of non–First Nations children, demonstrated that the First Nations cohort experienced significantly more physical pain–related diagnoses for 10 of the 13 pain indicators studied [7]. Further to this, while ear- and throat-related diagnoses were the 2 most diagnosed pain conditions for both groups, the First Nations cohort was significantly less likely than the non–First Nations cohort to visit a specialist for these same conditions, demonstrating inequities in diagnosis and treatment [7]. Additionally, these data showed that children (0 - 9 years) who had a physical pain diagnosis were more likely to have a mental health diagnosis in adolescence, a finding only evident in the non–First Nation cohort who were seemingly able to access mental health care [7].

Cultural variances in pain expression have also been found between Indigenous and non-Indigenous children and youths [8,9]. Latimer et al [8] concluded that Mi'kmaq children were stoic and hid their pain. Children reported that they would not cry when enduring pain and preferred instead to be "brave" or "tough it out" [8]. These findings led researchers to conclude that health clinicians may not have ways to accurately assess or document the level of pain a child or youth is experiencing. Similar findings of stoicism were reported in Indigenous peoples in Australia, with participants often being described as quiet about their pain and not reporting pain due to higher pain tolerance, fear of Western medicine, or intercultural communication difficulties [10]. Clinicians are trained to use numerical or face pain scales to assess pain; however, only minimal research has been done to determine if these self-report pain scales are culturally appropriate for use with Indigenous children, and indeed to our knowledge, none in First Nations children [11,12]. Community members reported that numerical pain rating assessment scales are confusing, and it is difficult to attach hurt or pain to one measure such as a face or number [9]. The evidence suggests that there is a clinical practice gap in assisting Indigenous youths to convey their pain to health providers. This study endeavored to find a culturally appropriate mechanism to support youths in sharing their pain and to address reports of a lack of clinician's ability to recognize pain, which creates a barrier to appropriate health care [13]. In response to these needs, the Kids Hurt App [14] was cocreated with Indigenous and clinical partners. It was created with the notion that Indigenous peoples have many effective ways of communicating knowledge that include visual images and storytelling. This led the developers to explore how both Indigenous and Western ways could be used to develop a better assessment tool that puts the technology in the hands of the youths. The aim of this qualitative study was to test the usability and feasibility of the Kids Hurt App with First Nations youths and clinicians working with youths.

Background

Two-Eyed Seeing

The project used a Two-Eyed Seeing (TES) approach, a term coined by Elders Murdena and Albert Marshall, which recognizes the benefit of seeing Indigenous knowledge and experience from one eye while also seeing strengths from another eye's perspective that is different from one's own [15]. The Western perspective is most typically how health care is currently delivered. When both eyes are used, the ultimate benefit is achieved by effectively enhancing the health of Indigenous peoples through the practical sharing of knowledge [16].

Digital Health Interventions for Indigenous Children and Youths

Although there is a growing body of literature on digital health and app-based interventions to manage pain in children and youths, research on the development and usability of app-based interventions for Indigenous peoples is limited. In a review of digital health solutions for Indigenous mental well-being across Canada, the United States, Australia, and New Zealand, Hensel et al [17] indicated that co-designed app-based interventions had been shown to reduce anxiety and depression symptoms and contribute to well-being. The Aboriginal and Islander Mental Health Initiative for Youth App, or AIMhi-Y, was developed in consultation with Indigenous youths from Australia and the Torres Strait Islands [18]. The youths identified apps as a potential way to mitigate barriers to accessing help and highlighted the need for a strength-based approach [18,19]. A randomized controlled trial found that iBobbly, a suicide prevention app developed in consultation with Indigenous youths living in Australia, had only a minimal, nonsignificant impact on psychological well-being outcomes, although qualitative data indicated the app was helpful and that construct validity issues may have been the reason why there was not a significant impact [20]. To date, there is no research on the development or usability of apps for Indigenous children and youths in Canada [21].

Development of the Kids Hurt App

The Kids Hurt App was developed by the Aboriginal Children's Hurt and Healing Initiative for children aged 10 - 19 years. Funding for the app development and testing was received from the Canadian Institutes of Health Research (FRN# 162455 & SCA-145102), Indigenous Health Nursing Research Chair, Dalhousie University (FRN# 167603), and IWK Health Foundation. The Aboriginal Children's Hurt and Healing Initiative is a broad partnership, consisting of Indigenous community leaders, clinicians, elders, youths, researchers from universities, and health delivery systems. In keeping with Ownership, Control, Access and Possession (OCAP) principles

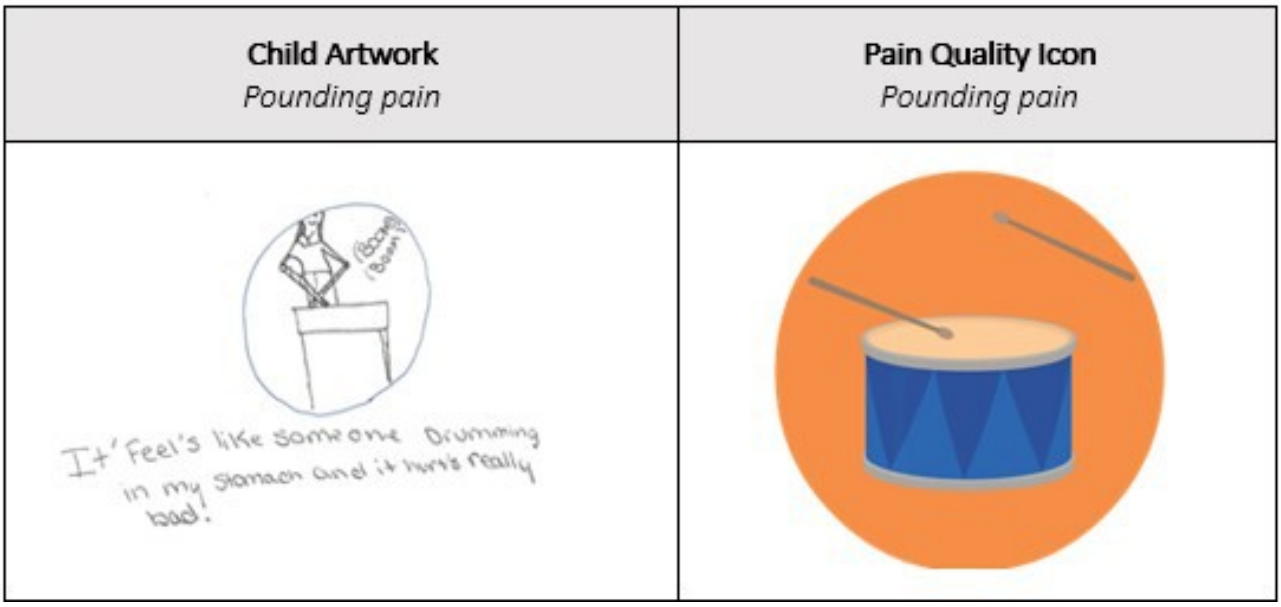
[22], the app is owned by communities involved in this research study. OCAP is a registered trademark of the First Nations Information Governance Centre and describes the principles of ownership, control, access, and possession to ensure that First Nations have control over data processes including ownership and how their information is used. The app was developed to facilitate pain communication between clinicians and Indigenous youths in a culturally safe manner. It is not meant to replace the history-taking step of a health care interaction but rather serves to augment the quality of communication exchange by giving youths an opportunity to share their pain story in a comfortable, culturally safe way.

The app was developed using TES to ensure both Indigenous and Western concepts and content were included. In total, 16 "screens" or interfaces were developed. Multimedia Appendix 1 provides an example of the final 16 screens developed. Previous versions of these screens were used throughout various stages of testing.

The first app screen asks users their gender and age (see Screen 1 in Multimedia Appendix 1), with the subsequent screen providing them with the option to create an avatar or choose a predesigned one (see Screen 2-3 in Multimedia Appendix 1). First Nations youths were involved in the development of the avatar designs through initial content validity testing to ensure that visual representation was present. Once an avatar is chosen, it is prompted to answer a series of responses to physical and emotional pain prompts. The inclusion of both physical and emotional pain prompts provides an example of how TES [15] was used throughout the development of the app. Earlier design ideas for the app had only physical pain prompts; however, community feedback suggested that incorporating both physical and emotional pain prompts would better represent the Indigenous holistic view of health, which involves "a healthy balance of 4 elements or aspects of wellness: physical, emotional, mental, and spiritual" [23]. While this version of the Kids Hurt App does not incorporate all 4 wellness aspects, emotional and physical pain were chosen because they were the most reported types of pain in previous research [7]. Future app iterations are expected to include spiritual and mental dimensions. There is also evidence that physical and emotional pain are interconnected, and the app presents an opportunity to highlight this for both youths and clinicians.

Additional incorporation of TES [15] involved the use of the International Association for the Study of Pain assessment tool guidelines [24] in the development, ensuring three key assessment features were embedded into the app: (1) pain location (see Screen 4-6 in Multimedia Appendix 1), (2) pain quality (see Screen 7 in Multimedia Appendix 1, and (3) pain intensity (see Screen 8 in Multimedia Appendix 1). Pain quality icons were conceptualized based on previous art-based research with Indigenous youths [8,9] as well as the Iconic Pain Assessment Tool [25] and the Pain Quality Icon (QuILT) [26] as illustrated in Figure 1.

Figure 1. Pain quality icon development.



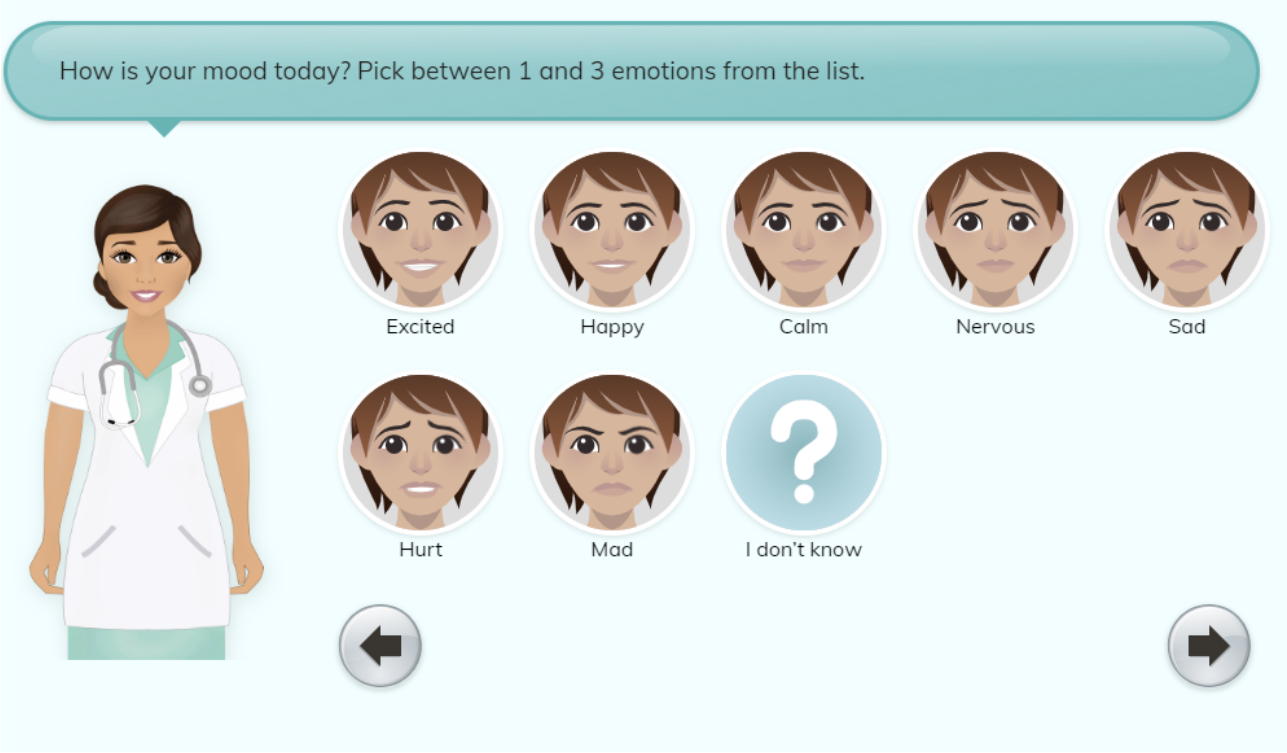
Pain intensity was initially captured using 2 scales: a 10-point "face scale," mimicking the Faces Pain Scale—Revised (FPS-R) [27], and a newly developed "Jar of Hurt," building on the concept of the "Pieces of Hurt" tool [28] and respecting the nation's language and word for "pain" which is "hurt" [8]. A jar was used as a common, everyday item that would provide the imagery of filling something up with pain or hurt. In addition, details of *where* users were when they got hurt (ie, school) and *what* they were doing (ie, playing) were also included with options based on common places of injury noted in the First Nations Regional Health Survey [29].The inclusion of the *where* and *what* aspects of the app allowed the incorporation of storytelling (see Screen 9-10 in Multimedia Appendix 1), an approach with significant cultural ties for Indigenous populations [30].

After completing questions regarding the user's physical pain, users are then prompted to provide information on their emotional pain through a series of similar screens (see Screen 11-14 in Multimedia Appendix 1). Emotional pain icons were initially chosen based on 22 common emotions identified from

20 validated mood apps (Multimedia Appendix 2). Through the initial content validity testing with First Nations youths, these 22 emotions were narrowed down to 8, which were identified as most likely to be used to describe their emotional mood. The resulting 8 emotions, both "negative" and "positive," provided a range of options and avoided bias in responses (Figure 2). Emotional pain questions focused on users' current emotional state (happy, sad, and anxious), intensity of those emotions (using the 10-point Jar of Hurt), and any physical symptoms that may be associated with their emotional pain (ie, chest tightness).

Upon answering both the physical and emotional pain questions, information is presented on a final storyboard screen (see Screen 15 in Multimedia Appendix 1).

Figure 2. Emotional pain icons.



Methods

Study Design

Once an initial version of the app was finalized three types of testing took place: (1) low-fidelity testing, (2) high-fidelity testing, and (3) clinical feasibility testing. This testing or development strategy mirrored the process published by Stinson et al [31].

Ethical Considerations

Prior to project initiation, the study was approved by both an Indigenous and tertiary pediatric hospital research ethics board (Mi'kmaw Ethics Watch & IWK Health Research Ethics Board; #1017831). Informed consent was obtained from all participants or their parents or guardians prior to taking part in study procedures. Youth participants received a CAD $10 (US $6.98) gift card after completing the testing. Clinician participants were not compensated. Data were deidentified immediately following data collection sessions.

Participants

First Nation youth participants were recruited through social media platforms and community organizations or contacts to complete low-fidelity, high-fidelity, and clinical feasibility testing. Clinicians recruited from a First Nation Health Centre and a pediatric health center unit specializing in emotional or physical pain–related services (rheumatology, pain clinic, etc) also completed high-fidelity testing. Eligibility criteria for youth participants included (1) self-identifying as Indigenous, (2) residing in a First Nation community, and (3) being between 10 and 19 years of age.

Data Collection

Overview

The study took place over a 5-year period between April 2017 and March 2022. Significant delays were experienced due to the COVID-19 pandemic. All youth data collection sessions were completed within the youth's Indigenous community to ensure that local support resources were available if a youth felt distressed. A trained mental health clinician was available as a resource for each session, and smudging was offered for youths who took part in the clinical feasibility sessions.

Low-Fidelity Testing

Low-fidelity testing is a qualitative usability approach where youths were able to view a paper-based version of the app. Youths completed a demographic survey prior to reviewing the app screens. In total, 3 rounds of low-fidelity testing took place. During each round, youths were shown paper-based screenshots of the Kids Hurt App and asked what they liked and disliked about the design through a series of semistructured questions.

Sessions were audio-recorded and transcribed for clarity and accuracy. After each round of testing, feedback was reviewed and integrated into the app. Screenshots of the updated version of the app were then shown to the next round of participating youths until no further changes were suggested based on the methods of Stinson et al [31].

High-Fidelity Testing

High-fidelity testing is similar to low-fidelity testing; however, participants were able to review a fully functioning version of the app rather than paper-based screenshots. Youths completed a demographic survey prior to using the app, and all sessions were audio-recorded and transcribed. Two rounds of high-fidelity testing were completed with participating youths to determine functionality and acceptability and to allow for any suggested revisions to be shared. The app was viewed on an iPad, and the youths worked through the app in the presence of the research assistant. At the start of each session, youths were asked to complete the app by themselves and to talk out loud about any likes, dislikes, or difficulties. The length of time it took for youths to complete the app from start to finish was recorded. Then, they were asked to go through the app for the second time with the research assistant who asked them semistructured questions. Once finished, they were asked open-ended questions on their overall thoughts regarding likes, dislikes, ease of use, and whether they would use the app in a health care setting if it was available. Feedback from each round was incorporated until no further changes were suggested (2 rounds). Interviews with the youths were audio-recorded in each of the 2 rounds to maintain the integrity and accuracy of the findings.

After the youths completed the high-fidelity testing, clinicians were then invited to take part in high-fidelity testing. Clinicians were invited to use the app and then answer a series of questions either through a digital survey or during a focus group. Focus group sessions were transcribed to ensure data accuracy. Questions assessed the clarity and importance of the various app components and also allowed space for participants to provide suggestions on what they would like to change or add for each screen. Demographic questions were also asked to determine participant's profession and types of experience.

Clinical Feasibility Testing

Clinical feasibility testing was completed with youths to help determine if the app works well, if it is easy to use, and if youths found it relevant for use in their care. Youths were asked to work through the app with the research assistant present and then answer a series of questions in a digital survey format. Questions ranged from demographic questions to those regarding the app's usefulness as well as functionality. Once an initial 5 youths had completed their review of the app, changes were incorporated, and an additional 5 youths were invited to take part. After 2 rounds of clinical feasibility testing with very positive results and minimal changes, it was determined that a third round was not needed.

Data Analysis

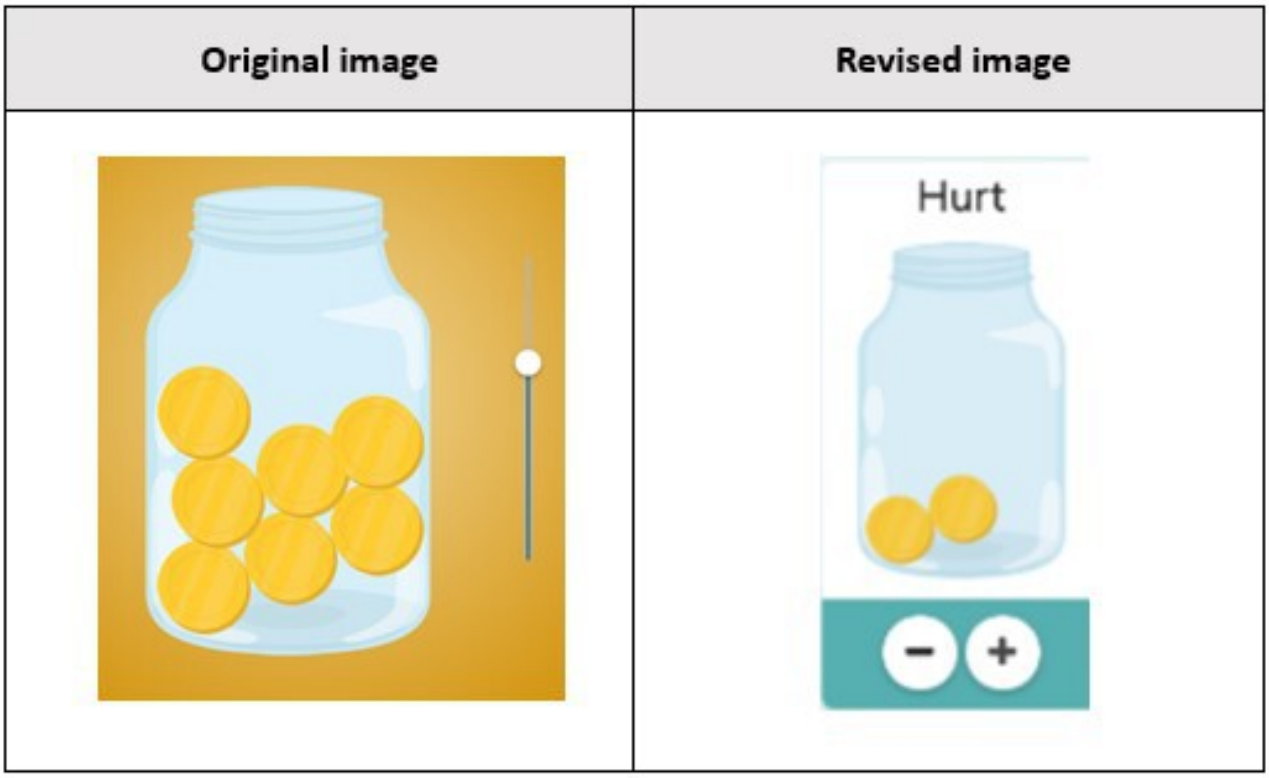
Demographic survey responses from all rounds of testing were analyzed using the SPSS Statistics (version 24; IBM Corp) database system to provide an overall breakdown of age and gender. Field notes and transcribed interviews were reviewed to determine any changes needed for subsequent rounds of testing. Revisions associated with more large-scale app functionality that were unable to be incorporated between testing rounds, such as incorporating a feature to track pain over time, were added to a database for future recommendations.

Results

Low-Fidelity Testing

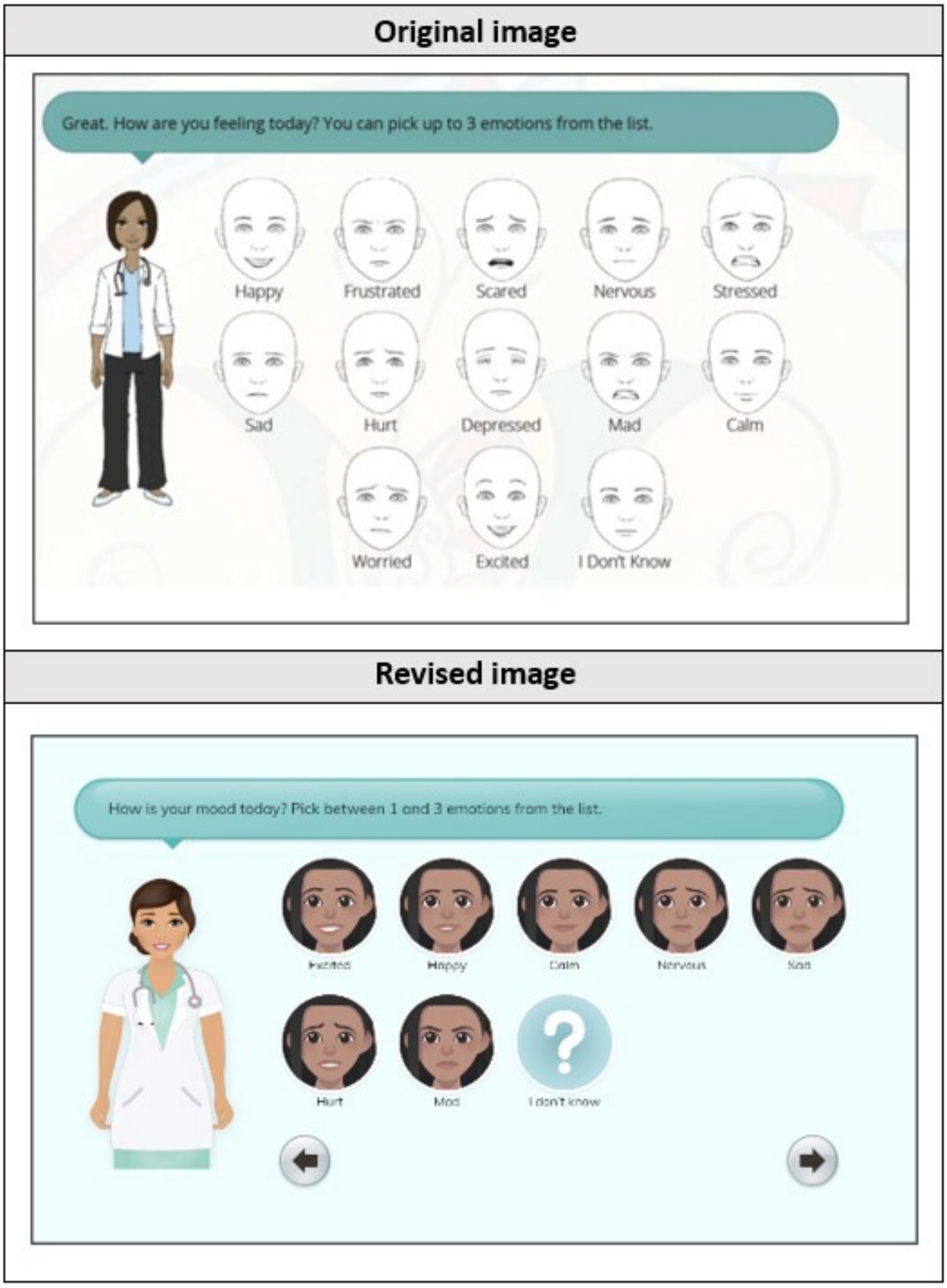
A total of 19 youths participated in 3 rounds of low-fidelity testing (round 1: 7 youths, round 2: 6 youths, and round 3: 6 youths). In total, 63% (n=12) of the participants were identified as female. The average age for all participants was 13 (SD 2.46) years. In the first round of low-fidelity testing, 1 participant suggested that the Jar of Hurt scale should use a plus or minus button to increase or decrease the amount of hurt rather than the original "slider" option (Figure 3). In subsequent rounds, youths were shown both the slider and the plus or minus option and asked their preference. The majority of youths (n=11) preferred the plus or minus option. The directions for describing how to use the Jar of Hurt were also identified during low-fidelity testing as needing clarification. A total of 6 of the 19 youths identified this issue. "I don't know, it's a little confusing. It sounds complicated" (Female, 14 years). The language was simplified, and it was agreed that it would be further reviewed in the high-fidelity rounds.

Figure 3. Jar of Hurt—plus or minus button addition.



The final noted change from low-fidelity testing was an overall dislike of the way the emotion face icons looked. One youth described the faces as “creepy.” To address this, the facial icons were redesigned to look like the user’s avatar rather than a generic face as illustrated in Figure 4.

Figure 4. Emotion faces—revised. Original image depicts 10 emotions rather than the finalized 8 emotions, as this was an initial mock-up screenshot for testing purposes only.



High-Fidelity Testing

A total of 20 youths participated in the 2 rounds of high-fidelity testing (round 1: 10 youths and round 2: 10 youths). In total, 70% (n=14) of participants were female, and the average age for youth participants was 16 (SD 2.46) years. The average time it took youths to complete the app was 3.61 minutes (216.6 seconds; SD 48). All youth participants identified enjoying the app and agreed it would be helpful to them.

> *It was pretty helpful. I know when you go to your doctor, you don't really want to open up. Sometimes you're nervous. At least the app helps explain how you feel.* [Female, 17 years]

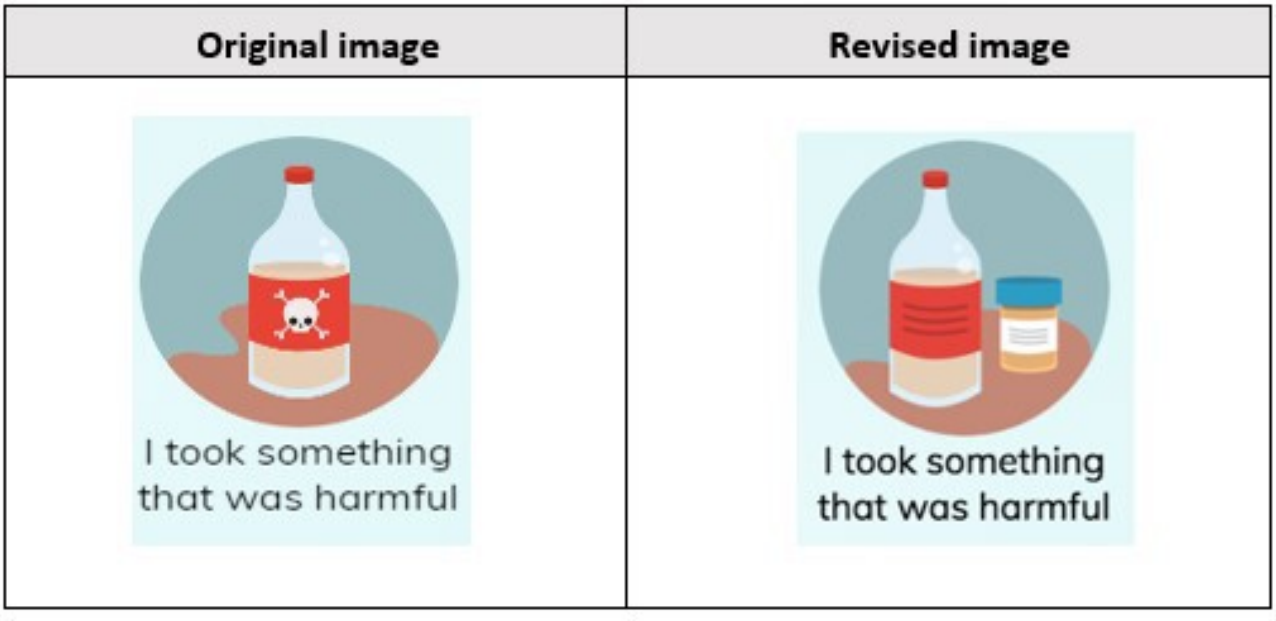
Confusion remained regarding the use of the Jar of Hurt with 2 of 10 youths in the first round of testing describing it as confusing. As a result, examples of an empty jar and a full jar were added beside the plus or minus buttons to provide a visual of how to increase or decrease the amount of pain or hurt (see Screen 7 and 8 in Multimedia Appendix 1). After this change was implemented, no confusion was noted in the remaining high-fidelity testing round.

Participants also suggested revisions to the "I took something harmful" icon in the "how you were hurt" section. Initially, this icon included a skull and crossbones to symbolize poison; however, youths suggested that something more representative of drugs or alcohol should be included.

> *I think this would be something chemical, based on the picture and skull thing. But I think something that can be added is maybe drugs or alcohol. Because the person may not always feel comfortable saying it but hitting it on the button without thinking and if it's right in front of you, they may be honest.* [Female, 18 years]

This icon was revised to better reflect youth participant's perception of which icons are most relevant to them (Figure 5).

Figure 5. "Harmful" icon changes.



A final revision resulting from the youth high-fidelity sessions was determining a preference for which scale to use to represent the quantity of pain. Youths were shown a scale with facial images similar to those in the FPS-R and superimposed in the avatar [27] (Figure 6) and the Jar of Hurt scale (Figure 7) and asked which they felt would best allow them to quantify their pain. The majority (n=12, 60%) preferred the Jar of Hurt scale. One youth shared their reason why:

> *If you had to choose one option, I would say the jar scale would be a lot better. Because maybe some people, they know how to hide it. Their pain. So, I'd say the jar scale is a lot better.* [Male, 16 years]

Figure 6. Faces Pain Scale—Revised expressions avatar adapted faces scale.

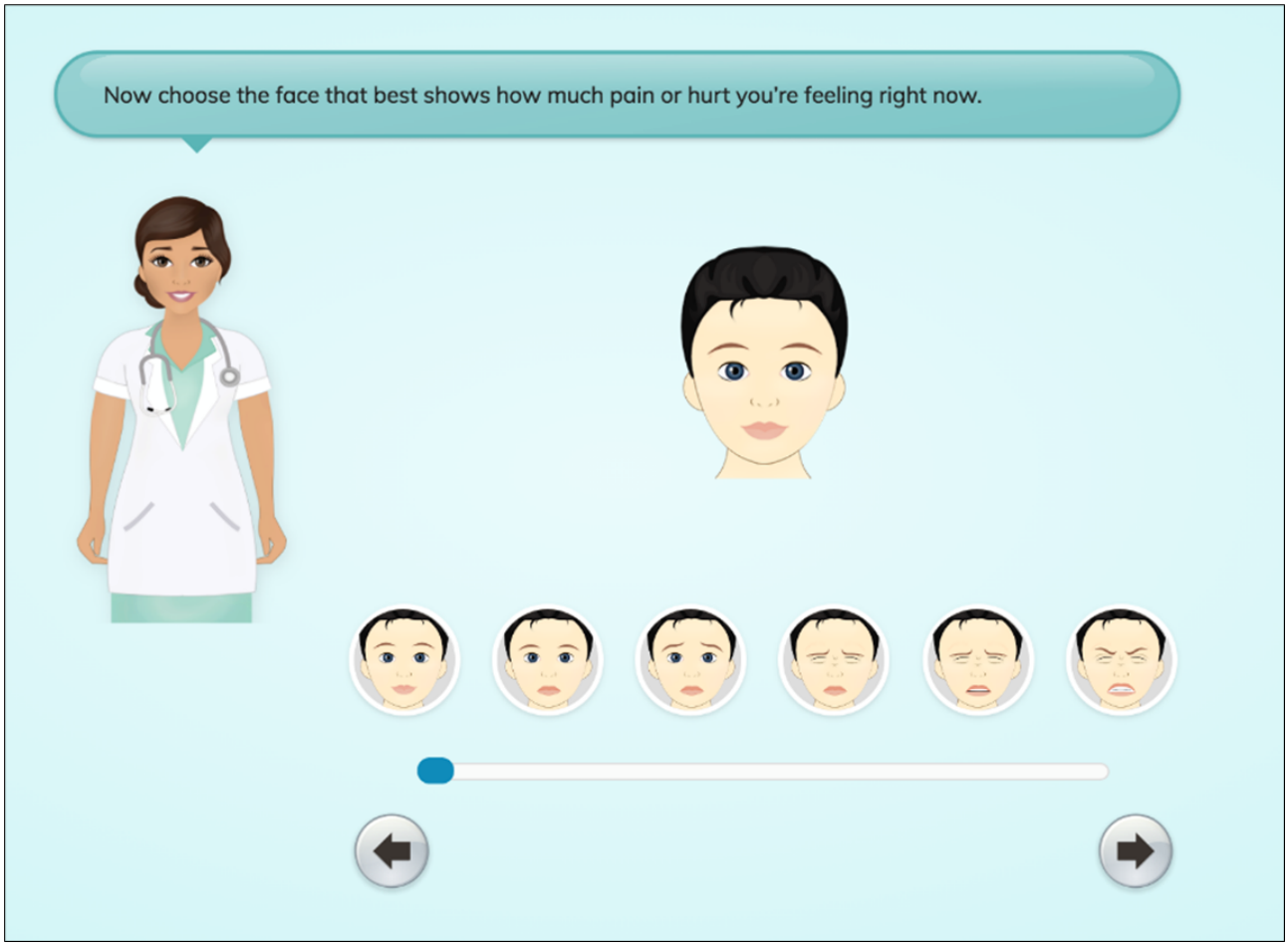
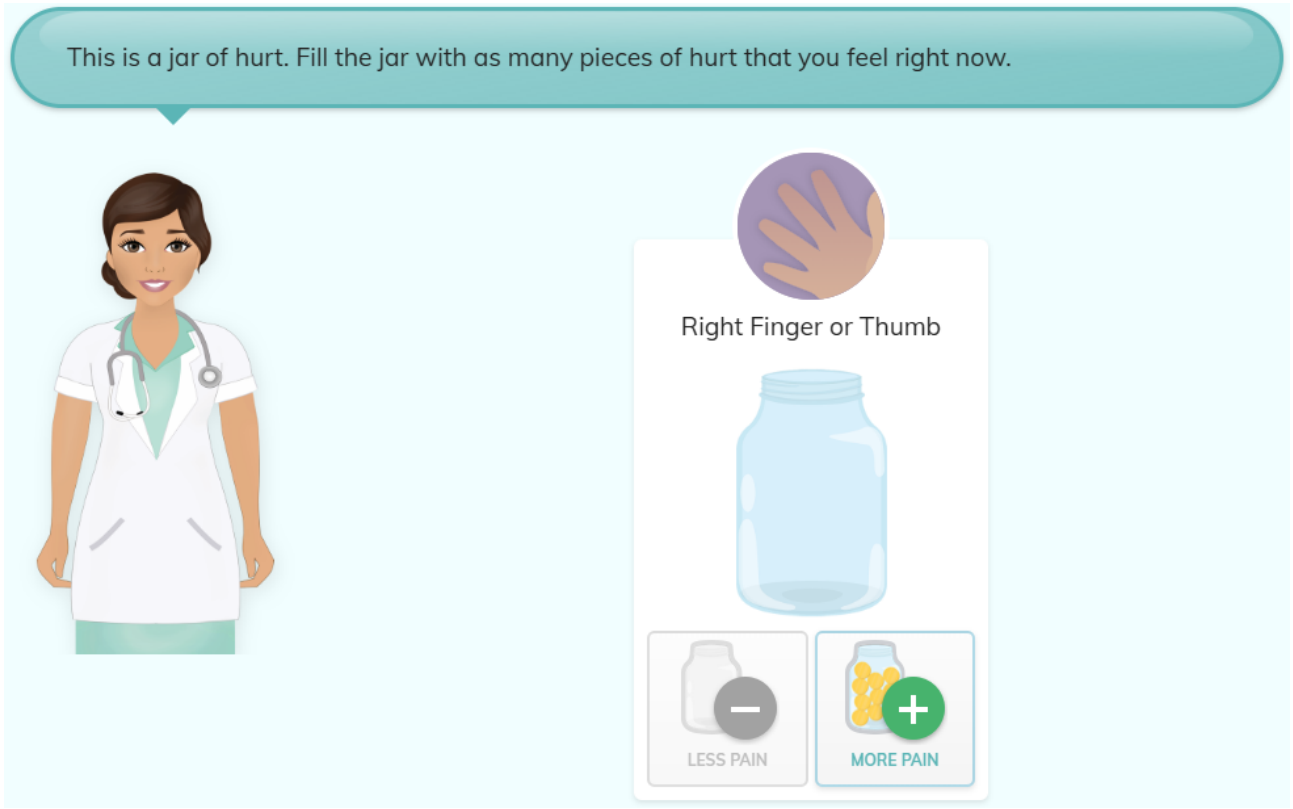


Figure 7. Jar of Hurt scale.



In addition to youths, a total of 15 clinicians participated in high-fidelity testing. In total, 10 clinicians completed this through a digital survey, and 5 participants took part in a focus group session. Most survey respondents were nurses (n=7, 70%), and focus group participants included physicians, physiotherapists, nurses, and psychologists.

In total, 14 of 15 clinicians found the app to be moderately or extremely useful. Like youths, clinicians were asked to identify their preference between the Jar of Hurt or the FPS-R adapted scale (Figures 6 and 7) through the survey questions. A total of 7 of the 9 clinician respondents identified a preference for the FPS-R adapted scale.

On 2 different screens, clinicians suggested adding an "other" option. For the quality of pain screen, 4 of 5 clinicians from the focus group felt that an "other" option should be included to allow users to add their own quality of pain descriptors. Similarly, for the emotional pain screen, all clinician participants suggested adding an "other" option to allow users the flexibility of adding any emotion they felt, even if not in the presented list. These additions were made in subsequent versions of the app.

Clinical Feasibility Testing

A total of 10 youths completed clinical feasibility testing (round 1: 5 and round 2: 5). Results from the clinical feasibility testing were very positive, with 80% (n=8) of respondents identifying that they found the app easy to use and 90% (n=9) saying they would use the app in a hospital or health care setting. No technical errors were found. Participants suggested that the Jar of Hurt method was more geared toward younger youths but also noted that they were happy to use the scale either way. Suggestions for change were mainly larger design changes identified as "future changes" including having a text or email function to send results to themselves or their care providers, having the ability to track changes in their pain or hurt over time, and including translation in their Indigenous language.

Discussion

Principal Findings

In this study of clinicians and First Nations youths, 3 types of testing occurred in the development of the Kids Hurt App: low-fidelity, high-fidelity, and clinical feasibility. For the 2 groups of participants, each subsequent round of testing reduced the app modifications. Overall changes were minor, but important observations were thought to increase the usefulness of the app for users.

Recent literature has identified that app-based interventions may remove barriers to accessing care for youths [18,32,33]. Apps are accessible and may mitigate confidential concerns and concerns about stigma. Smartphone-based apps may be particularly useful, as authors from one study found that smartphones are a mainstream technology and are an advantage within the health care system [34]. While existing research suggests that Western pain assessment tools are not appropriate for Indigenous peoples [10], this study demonstrates how an app can be developed collaboratively and engage youths to share their pain respecting both Indigenous and Western knowledge systems. All youths stated that they would use the Kids Hurt App if it was available in the health care setting. Youths reaffirmed that they often found it difficult to articulate their feelings and felt that the app could help them communicate their needs. This app allows them to share their pain and help facilitate enhanced communication.

One notable finding in this study was the process of determining which pain scale to use, the "Jar of Hurt" scale or the "face scale." In the initial development stages of the app, the "Jar of Hurt" concept was chosen based on previous research regarding the "Pieces of Hurt" tool [28]. While this research was completed with younger children than our target population, the concept of using objects rather than faces was appealing based on previous research indicating that Indigenous children are often stoic in their expression of pain and may hide their pain rather than displaying it outwardly through facial expressions [8]. Through high-fidelity testing, 60% (n=12) of youths preferred the "Jar of Hurt" scale over the "face scale," and while it was noted in clinical feasibility testing that the scale was more geared toward younger youths, participants identified being happy to use the tool, regardless of that. Clinician preference was strongly noted for the face scale likely because of its similarity to the FPS-R scale, a well-known pain measurement tool [27]. However, given that the app was developed to provide youths with a valid way to convey their pain and hurt and to develop an alternative to Western pain assessment tools, the "Jar of Hurt" was maintained as the preference for conveying pain intensity. This finding highlights the importance of taking the time to validate patient indicators of pain consistent with different age groups and populations.

Improved Health Care Experience

During a typical patient visit, a clinician obtains the patient's history, current pain experience, and future needs in a short amount of time [35]. This can be challenging if a patient finds it difficult to articulate their pain and hurt for a range of reasons, such as stoicism or power dynamic, further causing delay in timely and appropriate care [34,36]. In this study, participants identified the app as "easy to use" and were able to complete the app screens in an average of 3.61 minutes (216.6 seconds; SD 48). The youths' suggestions for app improvement, such as the Jar of Hurt being a better way to capture pain instead of a face that may hide pain, provide some indication that the Kids Hurt App may facilitate better communication, and consequently health care experience, by allowing Indigenous youths to convey their pain through an app. Other comments such as the youths who indicated pushing a button might be easier than talking could imply that the app offers a nonthreatening and safe way for youths to share their stories. Research has shown that youths have high technological literacy and are comfortable using eHealth apps [37,38]. In a generation where technology is fluent, the Kids Hurt App can give youths a voice in the health care system, in a medium they are comfortable with, to personally advocate for their physical and emotional pain care.

Limitations

At the time of testing, the app was not yet available in the youth participant nation's language. The current version of the app [14] is fully translated into Mi'kmaw. Communication between

the provider and patient should always be clear and effective, and having health information accessible in one's primary language is important [39]. Thus, we recommend that the app be available in the nation's primary language in any further testing. An additional limitation is that testing was only completed with 2 First Nation communities so results cannot be assumed to be accurate for all Indigenous nations.

Conclusions

Culturally safe care acknowledges the attitudes, knowledge, and behaviors of another's culture in a respectful manner. If a provider encounters an Indigenous patient, it would be important to understand the patient's unique historical legacies, ways of communicating, and any potential intergenerational trauma that may influence pain report [40]. Implications of creating and using an app that has been codeveloped by the end users (First Nations youths) increase the usefulness of the app and provide a space for the provider and patient relationship to be optimized for best care outcomes. The short app completion time is an indication of the minimal effort it takes to offer a culturally appropriate alternative to pain self-report and may reduce miscommunication and repeat visits for the same condition. Future research will include testing the app with non-Indigenous youths as well as in various health settings such as mental health, tertiary, primary care, and emergency departments. Further development of the app related to youth's self-report of effective strategies that reduce pain and hurt is planned. The Kids Hurt App is a culturally safe and evidence-based communication tool that may allow First Nations youths to comfortably share their pain and hurt stories. Each nation may be different in its communication and will need to review the app for usability and feasibility in its own communities.

Acknowledgments

The authors would like to thank the First Nations youths and the clinicians for taking the time to support the development and testing of the Kids Hurt App.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App screens.
[PDF File, 1154 KB - humanfactors_v12i1e48370_app1.pdf]

Multimedia Appendix 2
Mood app inventory.
[PDF File, 229 KB - humanfactors_v12i1e48370_app2.pdf]

References

1. Indigenous population continues to grow and is much younger than the non-Indigenous population, although the pace of growth has slowed. Statistics Canada. 2022. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/220921/dq220921a-eng.htm [accessed 2025-01-24]
2. Turpel-Lafond ME. In Plain Sight: addressing Indigenous-specific racism and discrimination in BC health care. 2020. URL: https://engage.gov.bc.ca/app/uploads/sites/613/2020/11/In-Plain-Sight-Full-Report-2020.pdf [accessed 2025-01-24]
3. Hamed S, Bradby H, Ahlberg BM, Thapar-Björkert S. Racism in healthcare: a scoping review. BMC Public Health 2022 May 16;22(1):988. [doi: 10.1186/s12889-022-13122-y] [Medline: 35578322]
4. Mahabir DF, O'Campo P, Lofters A, Shankardass K, Salmon C, Muntaner C. Experiences of everyday racism in Toronto's health care system: a concept mapping study. Int J Equity Health 2021 Mar 10;20(1):74. [doi: 10.1186/s12939-021-01410-9] [Medline: 33691682]
5. Goodman A, Fleming K, Markwick N, et al. "They treated me like crap and I know it was because I was Native": the healthcare experiences of Aboriginal peoples living in Vancouver's inner city. Soc Sci Med 2017 Apr;178:87-94. [doi: 10.1016/j.socscimed.2017.01.053] [Medline: 28214449]
6. McNally M, Martin D. First Nations, Inuit and Métis health: considerations for Canadian health leaders in the wake of the Truth and Reconciliation Commission of Canada report. Healthc Manage Forum 2017 Mar;30(2):117-122. [doi: 10.1177/0840470416680445] [Medline: 28929885]
7. Latimer M, Rudderham S, Lethbridge L, et al. Occurrence of and referral to specialists for pain-related diagnoses in First Nations and non-First Nations children and youth. CMAJ 2018 Dec 10;190(49):E1434-E1440. [doi: 10.1503/cmaj.180198] [Medline: 30530610]
8. Latimer M, Simandl D, Finley A, et al. Understanding the impact of the pain experience on Aboriginal children's wellbeing: viewing through a Two-Eyed Seeing lens. First Peoples Child Fam Rev 2014;9(1):22-37. [doi: 10.7202/1071791ar]
9. Latimer M, Young S, Dell C, Finley GA. Aboriginal children and physical pain: what do we know? IJIH 2012;9(1):7-14. [doi: 10.18357/ijih91201212389]

10. Strong J, Nielsen M, Williams M, Huggins J, Sussex R. Quiet about pain: experiences of Aboriginal people in two rural communities. Aust J Rural Health 2015 Jun;23(3):181-184. [doi: 10.1111/ajr.12185] [Medline: 25945587]
11. Jimenez N, Garroutte E, Kundu A, Morales L, Buchwald D. A review of the experience, epidemiology, and management of pain among American Indian. Alaska Native Aboriginal Can peoples J Pain 2011;12(5):511-522. [doi: 10.1016/j.jpain.2010.12.002] [Medline: 21330217]
12. Ellis JA, Ootoova A, Blouin R, et al. Establishing the psychometric properties and preferences for the Northern Pain Scale. Int J Circumpolar Health 2011 Jun;70(3):274-285. [doi: 10.3402/ijch.v70i3.17823] [Medline: 21703128]
13. Empathy, dignity, and respect: creating cultural safety for Aboriginal people in urban health care. Health Council of Canada. 2012. URL: https://learningcircle.ubc.ca/files/2014/05/Empathy-dignity-and-respect-Creating-cultural-safety-for-Aboriginal-people-in-urban-health-care.pdf [accessed 2025-01-24]
14. Kids Hurt App. ACHH Initiative. 2024. URL: https://www.kidshurtapp.com [accessed 2025-01-24]
15. Marshall A, Marshall M, Bartlett C. Two-Eyed Seeing in medicine. In: Greenwood M, Leeuw S, Lindsay N, Reading C, editors. Determinants of Indigenous Peoples' Health in Canada: Beyond the Social: Canadian Scholars Press; 2015:16-24.
16. Hovey RB, Delormier T, McComber AM, Lévesque L, Martin D. Enhancing Indigenous health promotion research through Two-Eyed Seeing: a hermeneutic relational process. Qual Health Res 2017 Jul;27(9):1278-1287. [doi: 10.1177/1049732317697948] [Medline: 28682710]
17. Hensel JM, Ellard K, Koltek M, Wilson G, Sareen J. Digital health solutions for Indigenous mental well-being. Curr Psychiatry Rep 2019 Jul 1;21(8):68. [doi: 10.1007/s11920-019-1056-6] [Medline: 31263971]
18. Povey J, Sweet M, Nagel T, et al. Drafting the Aboriginal and Islander Mental Health Initiative for Youth (AIMhi-Y) app: results of a formative mixed methods study. Internet Interv 2020 Sep;21:100318. [doi: 10.1016/j.invent.2020.100318] [Medline: 32477884]
19. Dingwall KM, Puszka S, Sweet M, Nagel T. "Like drawing into sand": acceptability, feasibility, and appropriateness of a new e‐mental health resource for service providers working with Aboriginal and Torres Strait Islander people. Aust Psychol 2015 Feb 1;50(1):60-69. [doi: 10.1111/ap.12100]
20. 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 Dec 1;7(12):e14296. [doi: 10.2196/14296] [Medline: 33258782]
21. Noronha N, Smith SJ, Martin Hill D, et al. The use of mobile applications to support Indigenous youth wellbeing in Canada. IJCAR 2021;8(1):124-134. [doi: 10.7202/1077724ar]
22. The First Nations principles of OCAP. First Nations Information Governance Centre. 2012. URL: https://fnigc.ca/ocap-training/ [accessed 2025-01-24]
23. An overview of Aboriginal health in Canada. National Collaborating Centre for Aboriginal Health. 2013. URL: https://www.ccnsa-nccah.ca/docs/context/FS-OverviewAbororiginalHealth-EN.pdf [accessed 2025-01-24]
24. Terminology. International Association for the Study of Pain. 2018. URL: https://www.iasp-pain.org/resources/terminology/ [accessed 2025-1-24]
25. McMahon E, Wilson-Pauwels L, Henry JL, Jenkinson J, Sutherland B, Brierley M. The Iconic Pain Assessment Tool: facilitating the translation of pain sensations and improving patient-physician dialogue. J Biocommun 2008;34(2):20-24 [FREE Full text]
26. Lalloo C, Kumbhare D, Stinson JN, Henry JL. Pain-QuILT: clinical feasibility of a web-based visual pain assessment tool in adults with chronic pain. J Med Internet Res 2014 May 12;16(5):e127. [doi: 10.2196/jmir.3292] [Medline: 24819478]
27. Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B. The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. Pain 2001 Aug;93(2):173-183. [doi: 10.1016/S0304-3959(01)00314-1] [Medline: 11427329]
28. Hester NK. The preoperational child's reaction to immunization. Nurs Res 1979;28(4):250-255. [Medline: 255924]
29. First Nations Regional Health Survey (RHS) 2008/10: national report on adults, youth and children living in First Nations. First Nations Information Governance Centre. 2012. URL: https://fnigc.ca/wp-content/uploads/2020/09/5eedd1ce8f5784a69126edda537dccfc_first_nations_regional_health_survey_rhs_2008-10_-_national_report_adult_2.pdf [accessed 2025-01-24]
30. Chan AS. Chapter 11: storytelling, culture, and indigenous methodology. In: Bainbridge A, Formenti L, West L, editors. Discourses, Dialogue and Diversity in Biographical Research 2021:170-185. [doi: 10.1163/9789004465916_012]
31. Stinson JN, Jibb LA, Nguyen C, et al. Development and testing of a multidimensional iPhone pain assessment application for adolescents with cancer. J Med Internet Res 2013 Mar 8;15(3):e51. [doi: 10.2196/jmir.2350] [Medline: 23475457]
32. Seko Y, Kidd S, Wiljer D, McKenzie K. Youth mental health interventions via mobile phones: a scoping review. Cyberpsychol Behav Soc Netw 2014 Sep;17(9):591-602. [doi: 10.1089/cyber.2014.0078] [Medline: 25007383]
33. Uhlhaas P, Torous J. Digital tools for youth mental health. NPJ Digit Med 2019;2:104. [doi: 10.1038/s41746-019-0181-2] [Medline: 31646184]
34. Suso-Ribera C, Mesas Á, Medel J, et al. Improving pain treatment with a smartphone app: study protocol for a randomized controlled trial. Trials 2018 Feb 27;19(1):145. [doi: 10.1186/s13063-018-2539-1] [Medline: 29482614]

35. Tai-Seale M, McGuire TG, Zhang W. Time allocation in primary care office visits. Health Serv Res 2007 Oct;42(5):1871-1894. [doi: 10.1111/j.1475-6773.2006.00689.x] [Medline: 17850524]
36. Pathak EB, Wieten SE, Wheldon CW. Stoic beliefs and health: development and preliminary validation of the Pathak-Wieten Stoicism Ideology Scale. BMJ Open 2017 Nov 14;7(11):e015137. [doi: 10.1136/bmjopen-2016-015137] [Medline: 29138193]
37. Thackeray R, Hunter M. Empowering youth: use of technology in advocacy to affect social change. J Comput Mediat Commun 2010;15(4):575-591. [doi: 10.1111/j.1083-6101.2009.01503.x]
38. Pretorius C, Coyle D. Young people's use of digital tools to support their mental health during Covid-19 restrictions. Front Digit Health 2021;3:763876. [doi: 10.3389/fdgth.2021.763876] [Medline: 34927133]
39. Molina RL, Kasper J. The power of language-concordant care: a call to action for medical schools. BMC Med Educ 2019 Nov 6;19(1):378. [doi: 10.1186/s12909-019-1807-4] [Medline: 31690300]
40. Dell EM, Firestone M, Smylie J, Vaillancourt S. Cultural safety and providing care to Aboriginal patients in the emergency department. CJEM 2016 Jul;18(4):301-305. [doi: 10.1017/cem.2015.100] [Medline: 26565079]

Abbreviations

FPS-R: Faces Pain Scale—Revised
OCAP: Ownership, Control, Access and Possession
TES: Two-Eyed Seeing

Edited by A Kushniruk, EM Borycki; submitted 20.04.23; peer-reviewed by D Dunsmuir, N Mungoli; revised version received 11.10.24; accepted 27.11.24; published 03.03.25.

Please cite as:
Francis K, Francis J, Latimer M, Gould H, Blackmore S, MacLeod E
Development and Testing of the Kids Hurt App, a Web-Based, Pain Self-Report App for First Nations Youths: Mixed Methods Study
JMIR Hum Factors 2025;12:e48370
URL: https://humanfactors.jmir.org/2025/1/e48370
doi:10.2196/48370

Understanding Experiences of Telehealth in Palliative Care: Photo Interview Study

Mahima Kalla[1], BE, PhD; Teresa O'Brien[1], BMedSci, PhD; Olivia Metcalf[1], PhD; Rashina Hoda[2], BSc, PhD; Xiao Chen[3], PhD; Andy Li[2], BCS, BEngEE (Hons); Catriona Parker[4], BAppSc, MAppSc, PhD; Michael Edward Franco[2,5,6], MBBS; Sam Georgy[7], BSc, BEng, MBA, MSc; Kit Huckvale[1], MBChB, BSc, MSc, PhD; Christopher Bain[8], MBBS, MInfoTech, PhD; Peter Poon[9,10], MBBS, PGDipPM, CHIA

[1]Centre for Digital Transformation of Health, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne Connect, 700 Swanston St, Carlton, Australia

[10]Department of Medicine, School of Clinical Sciences at Monash Health, Faculty of Medicine,Nursing & Health Sciences, Monash University, Melbourne, Australia

[2]Faculty of Information Technology, Monash University, Melbourne, Australia

[3]School of Information and Physical Sciences, College of Engineering, Science and Environment, University of Newcastle, Callaghan, Australia

[4]Transfusion Research Unit, Faculty of Medicine, Nursing & Health Sciences, Monash University, Clayton, Australia

[5]Faculty of Medicine, Nursing and Health Sciences, Monash University, Clayton, Australia

[6]St Vincent's Health, Melbourne, Australia

[7]Healthdirect Australia, Sydney, Australia

[8]Alliance for Digital Health At Monash, Faculty of Information Technology, Monash University, Clayton, Australia

[9]Supportive and Palliative Care Unit, Monash Health, Melbourne, Australia

Corresponding Author:
Mahima Kalla, BE, PhD
Centre for Digital Transformation of Health, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne Connect, 700 Swanston St, Carlton, Australia

Abstract

Background: It is widely accepted that the COVID-19 pandemic has accelerated the era of online health care delivery, including within community palliative care. This study was part of a larger project involving a collaboration between universities, health care services, government agencies, and software developers that sought to enhance an existing telehealth (video call) platform with additional features to improve both patient and health care professional (HCP) experience in a palliative care context.

Objective: The aim of this study was to understand palliative care patients' and HCPs' experiences of telehealth delivery in a palliative care context in Victoria, Australia. For the purposes of this study, telehealth included consultations by both video and telephone calls. By better understanding users' experiences and perceptions of telehealth, we hoped to determine users' preferences for new telehealth enhancement features.

Methods: A total of 6 health care professionals and 6 patients were recruited from a major tertiary hospital network's palliative care unit in Victoria, Australia. Participants were asked to generate 3 - 5 photographs depicting their telehealth experiences. These photographs were used as visual aids to prompt discussion during subsequent one-on-one interviews. Intertextual analysis was conducted to identify key themes.

Results: A total of 3 overarching themes emerged: comfort (or lack thereof) afforded by telehealth, connection considerations in telehealth, and care quality impacts of telehealth. Patients (n=6) described telehealth as supporting their physical and psychological comfort and maintaining connection with HCPs, yet there were specific situations where it failed to meet their needs or impacted care quality and delayed treatment. HCPs (n=6) recognized the benefit of telehealth for patients but reported several limitations of telehealth, in particular due to lack of physical examination opportunities. Participants indicated that 2 types of connection were imperative for effective telehealth delivery: technical connection (eg, good internet connectivity or clear phone line) and interpersonal connection (ie, good rapport and therapeutic alliance between the HCPs and patients). Often technical connection issues impeded the development of interpersonal connection between the HCPs and patients in telehealth.

Conclusions: The findings presented in this study combined with other co-design activities, which are outside the scope of this paper, indicated the potential value of a telehealth enhancement feature that generates patient-facing clinical consultation summaries. Our team has developed a video telehealth enhancement feature (or "add-on"), which will enable clinicians to distill key actionable advice and self-management guidance discussed during teleconsultations for a take-home summary document for patients. The add-on's prototype has also been subjected to an initial simulation study, which will be reported in a future publication.

(JMIR Hum Factors 2025;12:e53913) doi:10.2196/53913

KEYWORDS

consultation summary; digital scribe; qualitative research; telehealth; digital health; photo-elicitation; palliative care; photo interview; qualitative research; photographs; intertextual analysis

Introduction

In Australia, chronic illness continues to be a major challenge in health care [1] with more than three-quarters (78.6%) of Australians impacted by at least one chronic condition [2]. Many of these chronic conditions are life-limiting [3], and thus patients may benefit from palliative care [4]. The 2018 National Palliative Care Strategy recommends the approach of using palliative care as a "means to improve the quality of life of patients and their families affected with life-threatening illnesses through the prevention and relief of suffering by means of early identification, assessment, and treatment of pain and other problems" [5]. Studies have shown that palliative care improves quality of life by improving independence, providing symptom relief and pain control, and supporting the physical, emotional, and spiritual needs of patients, their carers, and loved ones [6-9]. Furthermore, palliative care is linked with a significant decrease in health care use and prolongment of survival [10].

Specialist palliative care in Australia is delivered using different models of care depending on the geographic location and available resources [11]. Many of these models of care were transformed during the COVID-19 pandemic to promote physical distancing [12,13]. One such approach for the delivery of specialist palliative care in Australia was a move to telehealth to limit hospital attendance and in-home visits [12]. The implementation of telehealth in palliative care in Australia demonstrated promise beyond the pandemic, with the potential to improve communication, facilitating more equitable access to palliative care (especially for those in rural and regional areas), avoiding lengthy travel to hospitals or clinics (particularly when patients are very unwell), and potentially assisting with timely symptom relief [14]. However, for continued sustainability and success of telehealth models, telehealth needs to adapt and evolve based on feedback from patients and health care professionals (HCPs) [15].

The importance of user involvement in the implementation of palliative care services has been recognized and prioritized by researchers to improve productivity, quality, and relevance of care [16]. The aim of this study was to understand palliative care patients' and HCPs' experiences of telehealth delivery in a palliative care context in Victoria, Australia. For the purposes of this study, telehealth included consultations by both video and telephone calls. This study was part of a larger project involving a collaboration between universities, health care services, government agencies, and software developers, which sought to enhance an existing telehealth (video call) platform with additional features to improve both patient and HCP experience. By better understanding users' experiences and perceptions of telehealth, we hoped to determine users' preferences for other telehealth enhancement features.

Methods

Participants

Palliative care professionals and patients were recruited from the Monash Health Supportive and Palliative Care Unit at a tertiary hospital network in Victoria, Australia. HCPs, including palliative care medical and nursing staff, and interpreters involved in the Unit's use of telehealth were invited to participate by email. The Unit has video telehealth palliative care clinics for oncology patients. In addition, during COVID-19 outbreak, the Unit provided a telehealth outreach service when in-person home visits were not possible.

Patients aged 18 years and older, with conversational English proficiency, currently using telehealth to access palliative care were invited through email invitations and postcards. A total of 6 patients and 6 HCPs participated in the study. All participants lived in Victoria and spoke English as their primary language. The patients' ages ranged between 54 and 61 years old. In addition, 4 of the 6 patient participants identified as female and 2 identified as male. The HCPs' ages ranged between 33 and 62 years old, with 5 identifying as female and 1 as male. The duration in their health care roles ranged from 6 months to 8 years. Their roles were 'doctor' (n=3), 'interpreter' (n=2), and 'nursing clinical coordinator' (n=1). Our 6 patient participants have been deidentified as P009, P012, P013, P014, P015, and P016, and HCP participants as HCP001, HCP003, HCP 004, HCP005, HCP010, and HCP011.

Data Collection

The study used a qualitative photo interviewing technique [17,18]. Participants were requested to generate 3‐5 photographs based on certain prompts and share them with the research team. HCPs were asked to click photographs to depict their everyday work life in the context of telehealth models of care, what telehealth represents in their work, general experiences, and opportunities for improving current telehealth models. Similarly, patients were asked to generate photographs to illustrate their experience of receiving health care through telehealth, what telehealth represents for them as patients and their everyday lives, and opportunities for improving telehealth models of care. Multimedia Appendix 1 and Multimedia Appendix 2 shows the specific instructions provided to both HCPs and patients.

Participants emailed their photographs to a member of the research team. Subsequently, telephone interviews, lasting between 30 and 45 minutes, were conducted with each HCP and patient participant. During the interviews, participants were asked to elaborate on what each of the photos meant for them and how the photographs related to their experiences of telehealth. During the interviews, use of participant-generated photographs helped trigger participants' memories and invoke their deeper consciousness, values, attitudes, and personal

meanings of telehealth [19-21]. The visual aids also helped the researcher and participant reach a shared understanding of the topic [22]. All the interviews were audio-recorded and transcribed verbatim by a professional transcribing company. The interview transcripts were deidentified, and photos blurred to protect participants' identities. The anonymized transcripts were imported into Microsoft Word and the photos inserted at the point where they were discussed during the interviews, allowing for an intertextual analysis in which the participants' photos and words could be viewed together [23].

Data Analysis

A multistep data analysis process was conducted, involving intertextual analysis wherein participants' words and photos were viewed together and analyzed using a 3-step process involving (1) preview, (2) review, and (3) cross-photo comparison [23]. In the first "preview" stage, the first author reviewed all the photos and identified recurring motifs. These motifs were recorded in a PowerPoint slide deck and shared with the research team at a meeting for feedback and further interpretation. Subsequently, a more in-depth "review" was conducted wherein one researcher coded all the interview transcripts with relevant themes and subthemes in more granular detail. The emergent codebook was shared and workshopped with the research team to further distill the emerging patterns.

The final stage in the data analysis process was "cross-photo comparison." During this phase, the research team discussed the similarities and differences within and across participant groups (eg, within patient participants' or within HCPs' accounts, or across patients' and HCPs' accounts). Often the patients' and HCPs' accounts provided contrasting experiences of the same facet of telehealth use (referred as "counter-examples"). For example, where patients valued the convenience of telehealth and being able to liaise with their doctors from anywhere, HCPs found patients' use of telehealth in certain places (eg, public spaces) disruptive and uncongenial to the therapeutic process. To organize the emergent themes, a storyboard was developed by the first author containing illustrative quotes and photographs from both patients and HCPs. This story board has been provided in Multimedia Appendix 3 . The storyboard was presented to the research team and subsequently, manuscript development commenced.

Ethical Considerations

Ethics approval for the study was obtained from the Monash Health Human Research Ethics Committee (project identification number 79681). All participants were provided plain language summaries of the research project, and what their participation would entail. All participants provided informed written consent. They also had the option to stop the interview at any point. Only approved members of the research team had access to the identifiable participant information. All interview transcripts were deidentified before the commencement of the data analysis process. Photos generated by participants showing any humans were also blurred to preserve participants' privacy.

Results

Overview

A total of 3 themes were identified: comfort (or lack thereof) afforded by telehealth, connection considerations in telehealth, and care quality impacts of telehealth. It should be noted that while these themes are presented as distinct categories, they are not necessarily mutually exclusive. In addition, in some cases the themes included counter-examples, as mentioned above.

Comfort (Or Lack Thereof) Afforded by Telehealth

For some participants, telehealth afforded them greater physical and psychological comfort. However, for some others, telehealth hampered their physical and psychological comfort. Therefore, this theme elucidates how telehealth use impacted patients' and HCPs' physical and psychological comfort.

Physical Comfort

Palliative care patients often experience physical discomfort, including pain and fatigue, which in turn impacts their ability to visit clinics for in-person consultations. Patient participants advised that they had arranged their homes in ways that minimized their discomfort. Telehealth enabled them to remain in this environment, unlike in-person appointments, which lead to physical exertion due to travel and wait times, with this issue compounded for rural patients.

Patient participant P015 captured this theme when discussing a photo (Figure 1) of the room where they spent most of their time. They reported being "in pain 24/7" and explained how telehealth helped minimize their discomfort:

> *...You can actually see how close my bed; my chair is to my office desk. And so, I just use the walking frame, I get to the chair very easily. I sit in my chair, I'm listening to a bit of music, or I'm just watching TV…I'll sit in the wait in the chair, waiting for the doctor to come in.* [P015]

Figure 1. Patient home office for telehealth consultations (P015).

By contrast, some HCPs associated telehealth with physical discomfort. HCPs working in the palliative care outreach service during the pandemic, as opposed to the Unit's dedicated telehealth clinics, reported that private rooms were not always available. Therefore, telehealth appointments were often conducted in shared workspaces, which could be noisy and distracting, impacting their ability to effectively conduct quality consultations with patients. For example, HCP001 included a photo of the shared office space (Figure 2) where they often conducted telehealth consultations and stated:

> *It gets very loud because we're [all] talking… So then anybody else talking on the phone, someone else has to talk up to talk over the top of that, and then anybody having a conversation with each other in the office has to talk up.* [HCP001]

Figure 2. Palliative care service shared office space (HCP001).

Psychological Comfort

Alongside physical comfort, telehealth provided patients with psychological comfort by reducing the stress associated with in-person appointments. Participants reported a variety of psychological stressors associated with in-person appointments, which were mitigated by using telehealth, for example, travel and wait times, risk of exposure to COVID-19 in hospitals, forgetting key questions for their doctors, and financial stress, particularly for rural patients who may need to make overnight accommodation arrangements for brief appointments. P013 provided a photo of a piggybank (Figure 3) and stated:

> *…going to appointments, car wear and tear, fuel for my daughter, parking fees, petrol... [With] telehealth, [I] don't have to spend that money… [the money saved could be spent on] all the bits and bobs that they keep chucking at me to take.* [P013]

Figure 3. Piggy bank (P013).

The psychological comfort associated with the freedom to stay in one's chosen location and receive health care from the comfort of their own homes is also illustrated in P014's photo (Figure 4) and quote as follows:

My bush family are my reason for living out here … This is my family. Because of telehealth, I get to stay home and enjoy them. You know, I don't lose a day or two being on the road away from them. I know it's only animals but for some people, you know, animals are our family. [P014]

Figure 4. Life in 'the bush' (P014).

For some outreach service HCPs, however, the sudden large-scale transition to telehealth in the aftermath of the COVID-19 outbreak, led to some psychological discomfort at their perceived inability to provide their usual standard of care. Given their professional training had primarily been in face-to-face care, they felt ill-equipped to efficaciously transition to online care. HCP005 provided a photo of themselves holding their stethoscope up to the camera (Figure 5), demonstrating the limitation of physical examination in telehealth consultations. HCP005 said that the return to more in-person consultations was "almost a relief" post–COVID-19 pandemic. They further elaborated:

> *…often to be thorough in your assessment, as a doctor, you've been trained [to] do your history, examination and do investigations. I guess telehealth takes away … what we're trained to do.* [HCP005]

Figure 5. Health care professional (HCP) uncomfortable about limitations of telehealth (HCP005).

Connection Considerations in Telehealth

Participants indicated that 2 types of connection were imperative for effective telehealth delivery: technical connection (eg, good internet connectivity or clear phone line) and interpersonal connection (ie, good rapport and therapeutic alliance between HCPs and patients). Often technical connection issues impeded the development of interpersonal connection between the HCPs

and patients in telehealth. This theme illustrates how technical connection and interpersonal connection impacted the participants' telehealth experiences.

Technical Connection

The technical aspects of connection included the skills required to use the technology, internet and mobile phone connectivity, and specific hardware and software concerns. HCP participants noted that their patients had first been introduced to the telehealth model of care during the pandemic with varying levels of technological proficiency. Patients with limited technological proficiency needed to build their confidence and skills over time, with support from HCPs, support staff, and family members. HCPs described situations where no one had checked whether or not a patient's computer had required microphone and video capabilities. Internet connectivity and bandwidth issues were also reported by both patients and HCPs. HCPs, in particular, due to their volume of telehealth calls, reported a lot of time lost to troubleshooting connectivity issues, which impacted the time available for clinical conversations. Often participants had to transition to telephone calls. For example, P012 shared an image of their landline home phone (Figure 6), describing it as follows:

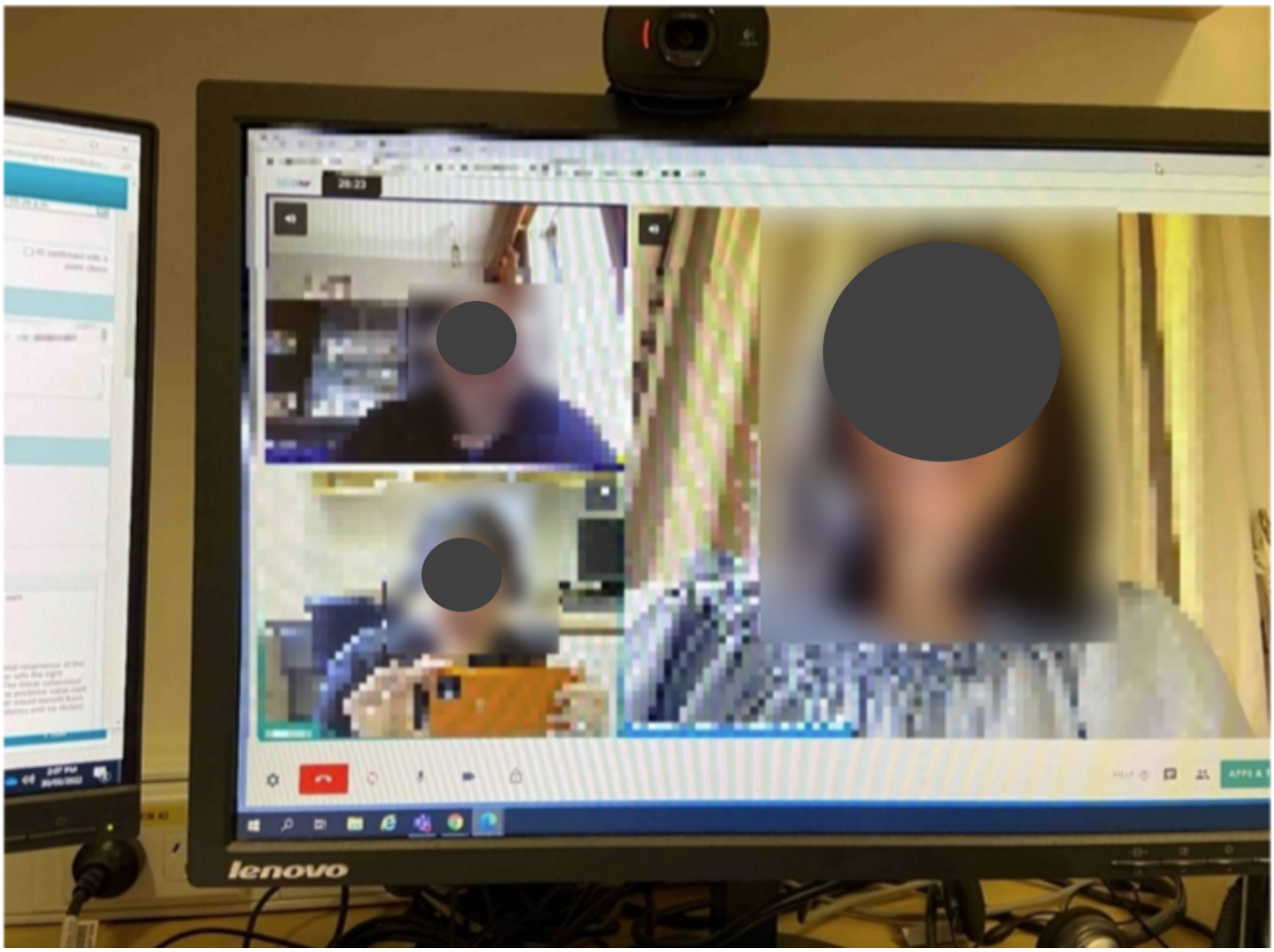
> *It is the most reliable tool in my house. It never drops out. Whereas the Internet … the internet on my computer can be slow or drop out or whatever at a telehealth appointment. And my mobile phone drops out, it's unreliable as well. But this phone is not unreliable. So if I lose a connection with any of the doctors, I know I can ring them on that phone … and it's not going to drop out. That's really important … So the landline is a good back up source for me for telehealth. I always have it sitting there in case something happens.* [P012]

Figure 6. Landline home phone as back-up (P012).

Nevertheless, participants also reported that in the absence of connectivity issues, and with the appropriate skill level to use, and availability of the right technology, telehealth can be very useful and seamless. For example, HCP003 illustrated a good experience of using telehealth involving both a patient and their carer in Figure 7, and recounted:

> *… essentially showing how we can make it work when everyone is capable of using telehealth. On the right … that's actually the patient's son … But it just shows how it can work. So … having patient caregiver in telehealth, it is all very possible. The video worked well the audio worked well ... there was no barrier there.* [HCP003]

Figure 7. Positive experience of telehealth (HCP003).

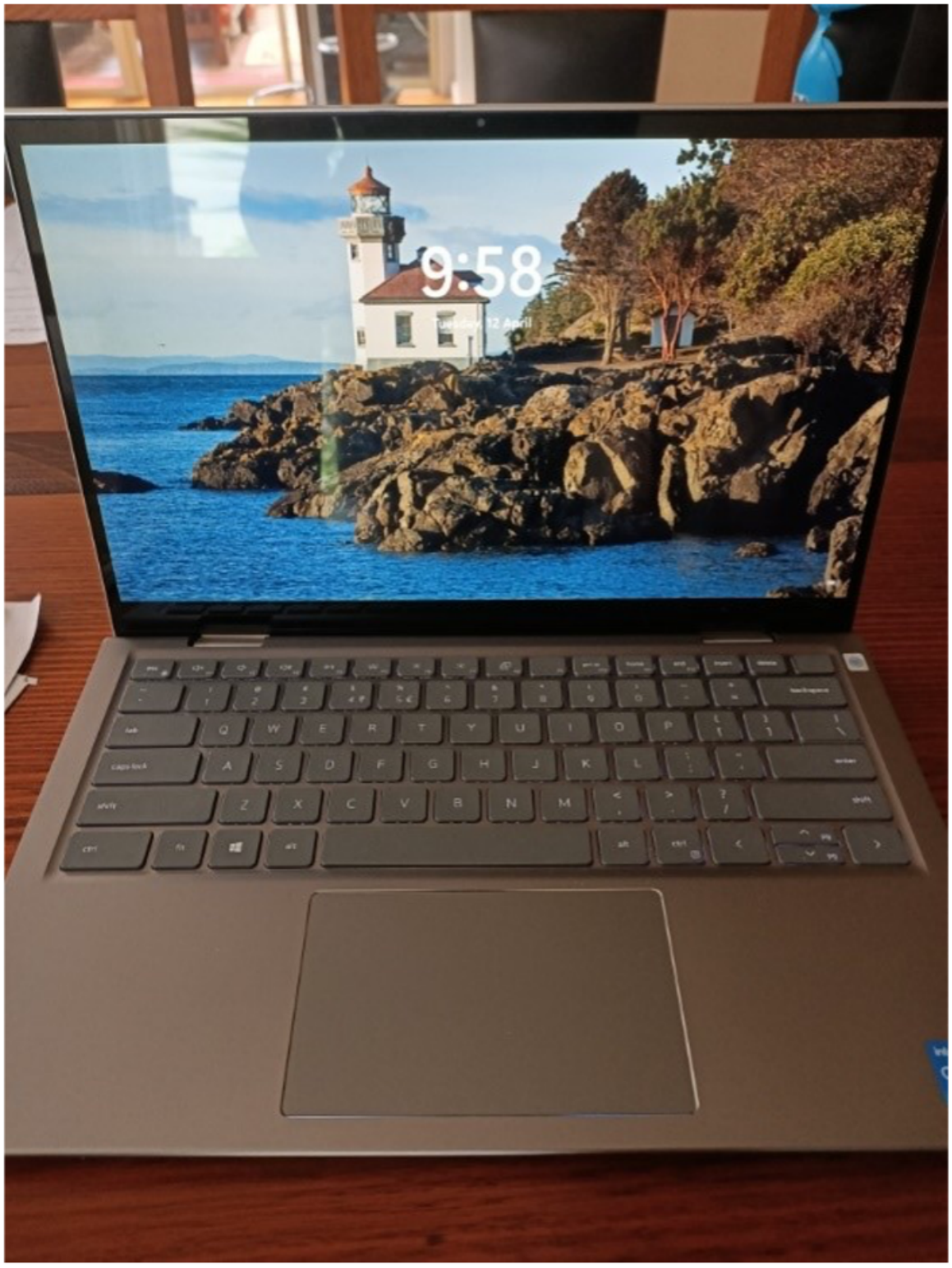


Interpersonal Connection

Participants expressed a general preference for video consultations over telephone consultations because of more opportunities for eye contact, and reading of facial expressions or other nonverbal cues. The notion of power dynamics between HCPs and patients also emerged, with some participants stating that conducting the consultations over telehealth made patients feel more relaxed and open, as opposed to feeling nervous in front of their HCP in person. P012 provided an image of their laptop (Figure 8), which they used for video consultations and stated:

> *I've noticed for me that I open up and talk more ... But when I'm in his rooms, I'm intimidated when I'm in person with him, face to face. I'm more conscious of myself and self-conscious. Whereas over the computer I'm not. I'm really confident and I just pour everything out.* [P012]

Figure 8. Patient's laptop used for video consultations makes them feel more relaxed (P012).



The type of technology used for video consultations also played a role in the quality of the clinical conversations. As previously mentioned, participants sometimes had to rely on backup options such as mobile phones or landlines for consultations due to technological failures, which also impacted the quality of their telehealth consultations. HCP011 shared a photo of an outreach service consultation that they had joined using their mobile phone, because the laptop and desktop were being used by other HCPs (Figure 9), and reflected:

> *The pictures are very small. So it just does not give you a realistic feeling of seeing a patient.* [HCP011]

Figure 9. Video consultation conducted on a mobile phone (HCP011).

In addition, participants generally agreed that periodic in-person consultations interspersed between their regular telehealth appointments were necessary to generate rapport between patients and HCPs. P015 had video telehealth oncology appointments every 6 - 8 weeks for over a year, despite which they said, "I've never met my oncologist," by which they meant that they had never met their oncologist in person. Even P012, who felt they were able to be more open through telehealth said:

> *the doctors should touch base with you specifically in person after a period of at least six months…. seeing them person to person sort of reconnects the relationship.* [P012]

In considering the benefits of in-person care, HCP010 said:

> *…we still want that relationship with the patient, we want to look them in the face and an important aspect of interpreting is body language. So even if you see them on the screen, [it] is not the same.* [HCP010]

HCP005 explained that some patients, "feel a bit suspicious or not completely trusting when you are not there [in person]."

Separately, HCP004 and HCP005 both also said that breaking bad news was more difficult through telehealth because of their inability to offer comfort through nonverbal communication like touch.

Care Quality Impacts of Telehealth

This final theme elucidates participants' perceptions about how telehealth impacts the quality of health care for patients.

Patient participants expressed mixed feelings regarding their satisfaction with the telehealth model of health care. While some participants perceived that the care they received through telehealth was satisfactory, others expressed concerns around issues such as lack of physical examination, impact on treatment timeliness, and communication.

Patients who described telehealth care positively viewed the crucial elements of care as unaffected. P014 explained that being able to share documents and show the HCP medications meant that:

> *[telehealth] doesn't stop the communication process in anyway…as [a] patient I'm not missing anything.* [P014]

Participant P014 was also able to monitor their blood pressure and oxygen saturation at home and share this information with their HCP. HCP004 saw telehealth as working well when there was another complementary team, for example, "hospital-in-the-home," which offered in-person care and provided clinical information to the palliative care patients. They also described the benefits of using telehealth to determine when an in-person visit was necessary:

> *it works best where there's opportunity for, or being able to identify that you can't rely on telehealth alone.* [HCP004]

Participants who believed that the lack of an in-person examination had compromised their care were more negative about telehealth. For example, P013 shared a photo of their arm in their lymphoedema sleeve (Figure 10) to describe the way telehealth had limited their care:

> *I was seeing her [my HCP] pretty much monthly on telehealth. As my arm would flare up, there was still not much she could do because we weren't allowed in … she finally got face to face with me, measured me up for the sleeve, so that was six months without proper diagnosis.* [P013]

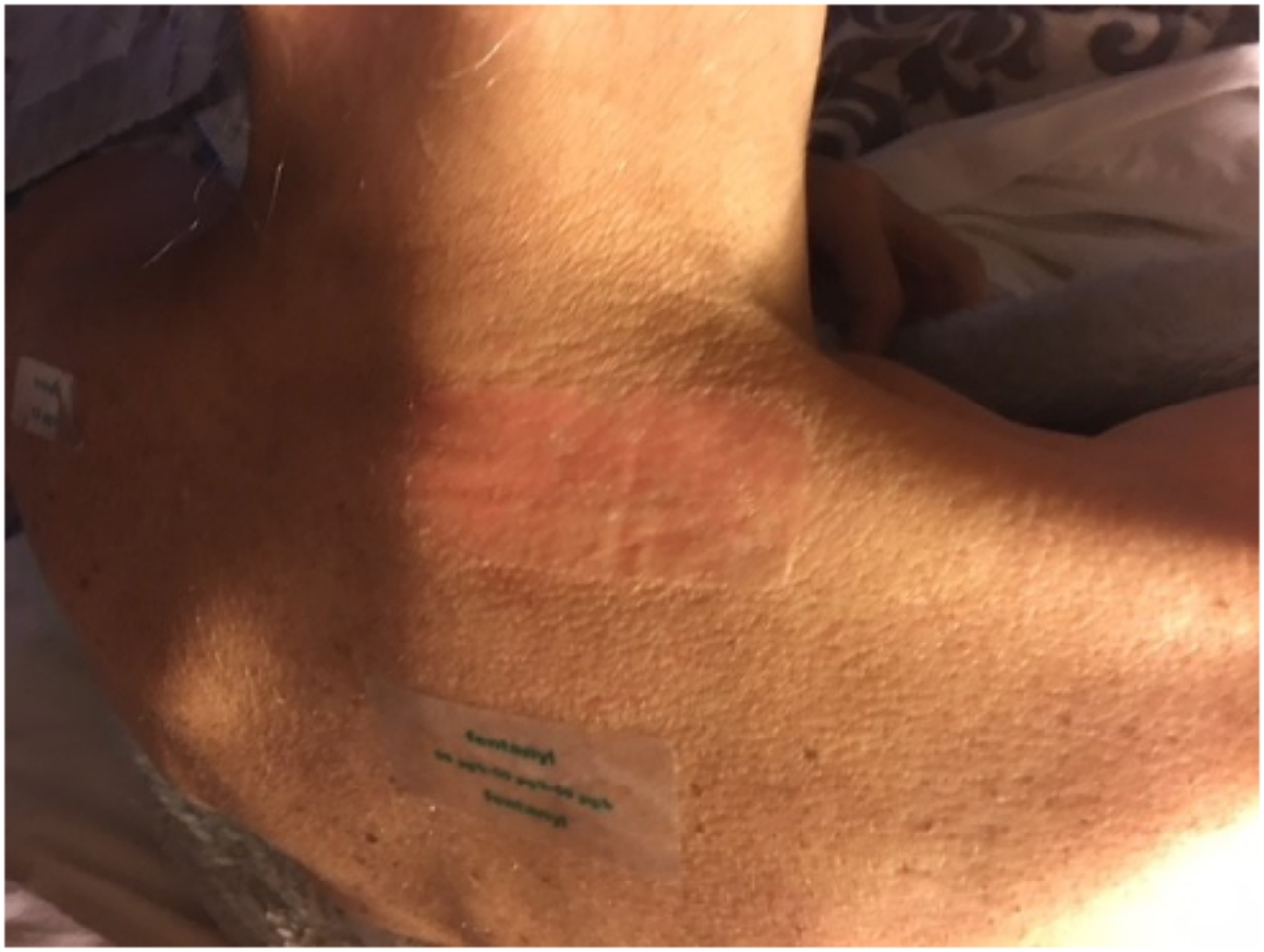
Figure 10. Lymphoedema sleeve (P013).

This participant perceived that their symptoms had not been taken as seriously or treated as effectively because the HCPs involved had not been able to examine them.

P016 used a photo of their skin (Figure 11) to discuss the implications of limited physical examination opportunities in telehealth:

> *I was trying to explain to the doctor what the problem was … he actually couldn't see me, what I was trying to explain he couldn't see … And we didn't really get to the bottom of what it was.* [P016]

Figure 11. Patient finding it difficult to show the health care professional (HCP) her skin redness concern (P016).



For this participant, in-person care was important not only in this specific instance, but also because:

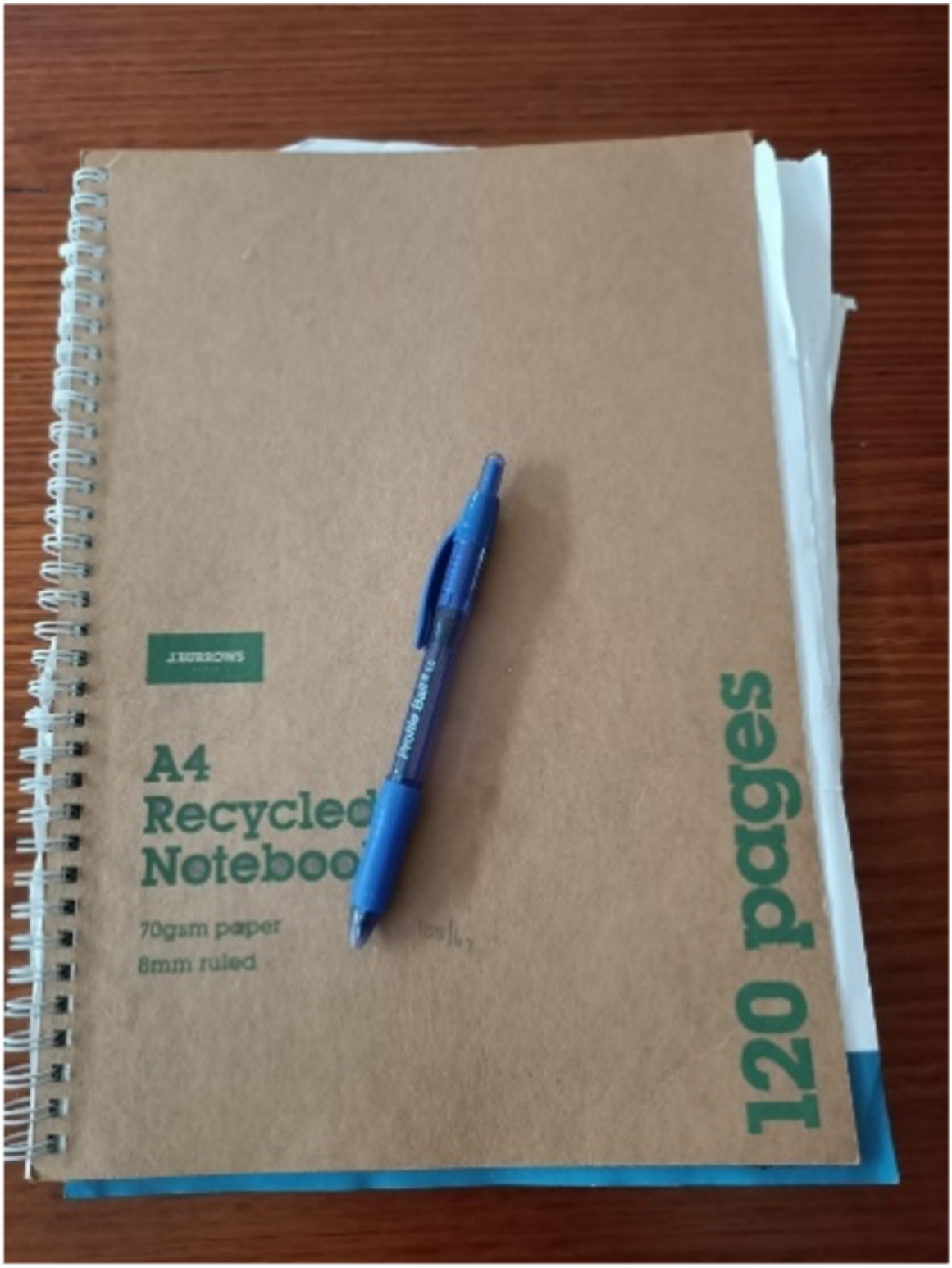
> *the doctor would pick up from the last time he saw you or your oncologist who last saw you [on] sort of things that you don't notice in yourself … I just feel as I'm getting more unwell, that I need to have more face-to-face appointments than telehealth appointments, but they tend to push you towards telehealth.* [P016]

HCPs also stated that telehealth sometimes resulted in less thorough assessments. HCP001 explained that this may be the result of poor technical connection, whereas HCP005 described it as a problem when the assessment required a third party to facilitate the telehealth consultation, such as staff in an aged care facility, whose other duties meant they did not have the requisite time to devote to the consultation.

On the other hand, a positive aspect of the telehealth model of care for patients was their ability to feel more in control of their self-care and self-management during telehealth consultations. For example, patient P012 took a photo of their notebook (Figure 12) to illustrate that they felt more comfortable bringing prewritten notes for discussion and scribing notes during a telehealth consultation than they would in person:

> *Notebook [is] very important ... Leading up to all my appointments I have, I make notes. And then when my appointment comes up, I cross stuff off … that's really important. For me, it's my notes. That way, I know I'm not going to forget anything. And whatever they say to me that's relevant or important, something I had to do or whatever, I'll write it down … I probably wouldn't do that at the doctor's … But online I can just do it on the side while I'm talking to them.* [P012]

Figure 12. Patient feels more in control of their ability to take notes before and during telehealth consultations (P012).



Similarly, another patient P009 recounted a similar experience of feeling more comfortable about their ability to capture important information before and during a telehealth consultation by having family members present with them from the comfort of their homes:

> *Part of the brain tumors is that I can't remember stuff afterwards. So I have to have someone there who's like my virtual notepad … they can answer things that I may have forgotten especially to do with medication because it seems to be the main one or [one of the] side effects … So it works well as in, they've [husband and/or daughter] got an area where they can be comfortable … I have a pad and pen here. But of course I have sometimes husband and/or daughter not only remembering stuff like that's said but going back and reminding me of stuff I need to tell the oncologist that I may have forgotten about.* [P009]

Discussion

Principal Findings

Our intertextual analysis yielded 3 key themes: comfort (or lack thereof) afforded by telehealth, connection considerations in telehealth, and care quality impacts of telehealth. Participants' relationship to telehealth was nuanced. Generally, participants acknowledged the myriad of benefits that telehealth brings, such as greater psychological and physical comfort for patients, especially those based in rural areas. However, there was also a general sense that telehealth should be combined with in-person care for optimal patient-HCP therapeutic alliance and better patient outcomes. It should be noted that this study's data collection commenced during the pandemic, amid nation-wide efforts to rapidly scale telehealth models of care across Australia. Eastman et al [15] sought to understand patients' and HCPs' experiences of telehealth in community palliative care. Their findings were congruent with ours, suggesting that telehealth is being increasingly used in palliative care and generally patients and HCPs are supportive of using this model of health care, combined with regular in-person touchpoints.

A salient finding from our study was the notion of greater power symmetry between patients and HCPs in telehealth compared with in-person consultations. Patients alluded to feeling more empowered, in control, and psychologically comfortable during teleconsultations. This translated into them feeling more confident about "pouring out" information in front of their HCPs, than they would face-to-face. Furthermore, the safety of a device's screen afforded patients the space to take notes during consultations. Having teleconsultations from the comfort of their homes where family members could also be present where needed, meant that patients felt more at ease. In some cases, family members also served the important function of remembering what had been discussed during consultations, which enhanced patients' quality of care.

Our team conducted additional co-design activities with consumers which are outside the scope of this article. In those activities too, participants had reported the importance of taking notes before, during, and after their telehealth appointments. Participants described forgetting important details from their appointments and relying on their family to attend consultations to remember what had happened during the appointment.

Literature suggests that 40% - 80% of medical information and recommendations provided by HCPs during clinical consultations is forgotten by patients almost immediately. The poor information recall may have downstream impacts for treatment adherence [24]. Issues of memory recall may be exacerbated in palliative care patients who experience complex illness symptomatology. There is some evidence to suggest that patient-facing clinical consultation recordings can be helpful for patients in general. A scoping review conducted by Tsulukidze et al [25] in 2014 suggested that patients place a high value on audio recordings of clinical consultations and benefit from subsequently listening to the consultation recordings.

The participants of our study reported in this paper and subsequent co-design activities indicated that having a summary of the consultation at the end of telehealth session would mean that this information is kept secure and in one place for patients. Memory and cognitive deficits are often experienced by palliative care patients. Therefore, a mechanism to record salient information discussed during consultations would be particularly useful for this patient population.

The literature on patient-facing consultation summaries generated during telehealth is currently limited. A non-systematic literature scan in Google Scholar on this subject yielded no relevant papers. The closest relevant literature appears to be about artificial intelligence (AI) enabled digital scribes or documentation systems, which can help automate clinical documentation tasks ordinarily conducted by humans [26-28]. However, the literature on patient-facing clinical summaries in telehealth (whether generated manually by HCPs or through AI-enabled algorithms) is not yet developed.

The paucity of literature on this subject combined with the participants' unanimous views emergent in our investigations, emphasized the need for innovating in the area of patient-facing clinical consultation summaries. Since the conduct of the study presented in this paper, our team of software developers has created a telehealth-enhancement add-on feature which seeks to generate patient-facing summaries of clinical consultations during video telehealth sessions. The prototype was developed in an agile manner, involving think-aloud testing activities with end users. Subsequently, the minimum viable product (MVP) was subjected to a simulation study to generate early-stage evidence about end-users' perceptions of the add-on's potential clinical usefulness. The findings of our think-aloud and simulation studies will be reported in future papers. However, based on participants' accounts presented in this paper, we expect that the provision of a patient-facing document that summarizes the telehealth consultation and distills key actionable clinical advice for patient self-management, might enhance patient experience and self-management.

Limitations

The limitations of this project include the small number of participants interviewed. The small sample size was partly due to challenges in recruiting palliative care patients. Patients' complex symptomatology and in some cases, limited life expectancy, impacts their ability to participate in research studies. In addition, as the data collection was conducted during the pandemic, this posed additional challenges for HCP recruitment. Nevertheless, our multimodal data collection approach involving the use of photographs and interviews combined with dual perspectives of patients and HCPs meant that we had a rich dataset to generate an in-depth understanding of users' experiences of telehealth.

In this instance, we did not collect information about, or pose eligibility restrictions regarding participants' ethnic and racial background, socioeconomic status, and level of experience with telehealth. This decision was driven by the small number of palliative care participant population available to participate in the study. Future telehealth research may include

contextualisation of emergent findings through the lens of one or more of these socio-demographic characteristics.

Conclusion

This study explored patients' and HCPs' experiences of telehealth in a palliative care context. A total of 3 themes emerged: comfort (or lack thereof) afforded by telehealth, connection considerations in telehealth, and care quality impacts of telehealth. The findings presented in this article combined with other co-design activities, which are outside the scope of this paper, indicated the potential value of a telehealth enhancement add-on feature that generates patient-facing clinical consultation summaries. Our team has developed a video telehealth enhancement feature, which will enable clinicians to distill key actionable advice and self-management guidance discussed during teleconsultations in a take-home summary document for patients. The add-on prototype has been subjected to a preliminary simulation study which will be reported in a future publication.

Acknowledgments

We thank Ms Roxanne Kritharidis who conducted all of the interviews, cleaned transcripts, and prepared the data for analysis. We thank Dr Jennifer Bibb who obtained ethics approvals, supported data collection, and participated in data analysis and writing of the final co-design report on which this work is based. We thank Dr Elton Lobo who contributed to analysis discussions and started early paper drafts. Thank you to Dr Kate Robins-Browne who led data analysis and drafted first versions of this paper. We also thank Prof Victoria Palmer who designed the photo interviewing study as part of the co-design stage for the early parts of this project, oversaw data collection and analysis, and contributed to the early draft of the paper. This research is supported by the Digital Health Cooperative Research Centre Limited (DHCRC), Monash University, Healthdirect Australia, Monash Health, University of Melbourne, and Victorian Department of Health. DHCRC is funded under the Commonwealth Government Cooperative Research Centres (CRC) Program. Finally, we thank all the participants of this study who generously contributed their time towards this research.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to the sensitive nature of the data and vulnerable nature of the population.

Generative artificial intelligence was not used for any part of the study reported in this paper nor for manuscript development.

Authors' Contributions

MK contributed to formal analysis, writing–original draft, writing–review and editing, and project administration. TOB contributed to formal analysis, writing–original draft, writing–review and editing, and project administration. CP and MEF contributed to writing–review and editing. CB, RH, and PP contributed to writing–review and editing and funding acquisition. XC, AL, and SG contributed to writing–review and editing and software. OM and KH contributed to writing–review and editing and project administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Photo instructions—health care professionals (HCPs).
[PDF File, 246 KB - humanfactors_v12i1e53913_app1.pdf]

Multimedia Appendix 2
Photo instructions—patients.
[PDF File, 202 KB - humanfactors_v12i1e53913_app2.pdf]

Multimedia Appendix 3
The 3 Cs storyboard.
[PDF File, 1357 KB - humanfactors_v12i1e53913_app3.pdf]

References

1. Harris MF, Harris-Roxas B, Knight AW. Care of patients with chronic disease: achievements in Australia over the past decade. Med J Aust 2018 Jul 16;209(2):55-57. [doi: 10.5694/mja18.00333] [Medline: 29996749]

2. Health conditions prevalence - key findings on selected long-term health conditions and prevalence in australia. Australian Bureau of Statistics. URL: https://www.abs.gov.au/statistics/health/health-conditions-and-risks/health-conditions-prevalence/latest-release [accessed 2022-10-19]
3. Deaths in Australia. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia/contents/leading-causes-of-death [accessed 2022-10-19]
4. Earp M, Cai P, Fong A, et al. Hospital-based acute care in the last 30 days of life among patients with chronic disease that received early, late or no specialist palliative care: a retrospective cohort study of eight chronic disease groups. BMJ Open 2021 Mar 24;11(3):e044196. [doi: 10.1136/bmjopen-2020-044196] [Medline: 33762238]
5. National Palliative Care Strategy 2018. Australian Commission on Safety and Quality in Health Care. URL: https://www.safetyandquality.gov.au/sites/default/files/2019-06/national_palliative_care_strategy_2018.pdf [accessed 2022-10-19]
6. Wajid M, Rajkumar E, Romate J, George AJ, Lakshmi R, Simha S. Why is hospice care important? An exploration of its benefits for patients with terminal cancer. BMC Palliat Care 2021 May 17;20(1):1-8. [doi: 10.1186/s12904-021-00757-8] [Medline: 34001076]
7. Otani H, Morita T, Igarashi N, Shima Y, Miyashita M. A nationwide survey of bereaved family members' perception of the place patients spent their final days: is the inpatient hospice like or unlike a home? Why? Palliat Med Rep 2020;1(1):174-178. [doi: 10.1089/pmr.2020.0058] [Medline: 34223474]
8. Currow DC, Agar MR, Phillips JL. Role of hospice care at the end of life for people with cancer. J Clin Oncol 2020 Mar 20;38(9):937-943. [doi: 10.1200/JCO.18.02235] [Medline: 32023154]
9. Shrank WH, Russell K, Emanuel EJ. Hospice carve-in-aligning benefits with patient and family needs. JAMA 2020 Jul 7;324(1):35-36. [doi: 10.1001/jama.2020.8459] [Medline: 32496529]
10. de Avila V, Paik JM, de Avila L, et al. Hospice care utilisation among elderly patients who died with hepatocellular carcinoma in the United States. JHEP Rep 2021 Apr;3(2):100236. [doi: 10.1016/j.jhepr.2021.100236] [Medline: 33748729]
11. Mitchell GK. Palliative care in Australia. Ochsner J 2011;11(4):334-337. [Medline: 22190884]
12. Pasanen L, Le Gautier R, Wong A, et al. Telehealth in outpatient delivery of palliative care: a qualitative study of patient and physician views. Pall Supp Care 2023 Dec;21(6):980-987. [doi: 10.1017/S1478951522000670]
13. Fitzgerald DA, Nunn K, Isaacs D. Consequences of physical distancing emanating from the COVID-19 pandemic: an Australian perspective. Paediatr Respir Rev 2020 Sep;35:25-30. [doi: 10.1016/j.prrv.2020.06.005] [Medline: 32690355]
14. Jiang B, Bills M, Poon P. Integrated telehealth-assisted home-based specialist palliative care in rural Australia: a feasibility study. J Telemed Telecare 2023 Jan;29(1):50-57. [doi: 10.1177/1357633X20966466] [Medline: 33079611]
15. Eastman P, Dowd A, White J, Carter J, Ely M. Telehealth: rapid adoption in community palliative care due to COVID-19: patient and professional evaluation. BMJ Support Palliat Care 2024 May;14(e1):e438-e442. [doi: 10.1136/bmjspcare-2021-002987]
16. Daveson BA, de Wolf-Linder S, Witt J, et al. Results of a transparent expert consultation on patient and public involvement in palliative care research. Palliat Med 2015 Dec;29(10):939-949. [doi: 10.1177/0269216315584875] [Medline: 25931336]
17. Burton A, Hughes M, Dempsey RC. Quality of life research: a case for combining photo-elicitation with interpretative phenomenological analysis. Qual Res Psychol 2017 Oct 2;14(4):375-393. [doi: 10.1080/14780887.2017.1322650]
18. Bugos E, Frasso R, FitzGerald E, True G, Adachi-Mejia AM, Cannuscio C. Practical guidance and ethical considerations for studies using photo-elicitation interviews. Prev Chronic Dis 2014 Oct 30;11:E189. [doi: 10.5888/pcd11.140216] [Medline: 25357257]
19. Glaw X, Inder K, Kable A, Hazelton M. Visual methodologies in qualitative research: autophotography and photo elicitation applied to mental health research. Int J Qual Methods 2017;16(1). [doi: 10.1177/1609406917748215]
20. Todd A, Holmes H, Pearson S, et al. "I don't think I'd be frightened if the statins went": a phenomenological qualitative study exploring medicines use in palliative care patients, carers and healthcare professionals. BMC Palliat Care 2016 Jan 29;15(1):1-7. [doi: 10.1186/s12904-016-0086-7] [Medline: 26822776]
21. Cederholm EA. The use of photo‐elicitation in tourism research – framing the backpacker experience. Scandinavian Journal of Hospitality and Tourism 2004 Jan;4(3):225-241. [doi: 10.1080/15022250410003870]
22. Pyyry N, Hilander M, Tani S. Photography and photo elicitation as visual methods. In: Creative Methods for Human Geographers: SAGE Publications; 2021:75-86. [doi: 10.4135/9781529739152.n6]
23. Palmer VJ, Furler J. A room with a view: a metaphor analysis of Vietnamese women's representations of living with depression using photo elicitation. Visual Studies 2018 Jul 3;33(3):251-263. [doi: 10.1080/1472586X.2018.1525309]
24. Kessels RPC. Patients' memory for medical information. J R Soc Med 2003 May;96(5):219-222. [doi: 10.1177/014107680309600504] [Medline: 12724430]
25. Tsulukidze M, Durand MA, Barr PJ, Mead T, Elwyn G. Providing recording of clinical consultation to patients - a highly valued but underutilized intervention: a scoping review. Patient Educ Couns 2014 Jun;95(3):297-304. [doi: 10.1016/j.pec.2014.02.007] [Medline: 24630697]
26. Coiera E, Kocaballi B, Halamka J, Laranjo L. The digital scribe. NPJ Digit Med 2018;1(1):58. [doi: 10.1038/s41746-018-0066-9] [Medline: 31304337]
27. Ghatnekar S, Faletsky A, Nambudiri VE. Digital scribe utility and barriers to implementation in clinical practice: a scoping review. Health Technol (Berl) 2021;11(4):803-809. [doi: 10.1007/s12553-021-00568-0] [Medline: 34094806]

28. van Buchem MM, Boosman H, Bauer MP, Kant IMJ, Cammel SA, Steyerberg EW. The digital scribe in clinical practice: a scoping review and research agenda. NPJ Digit Med 2021 Mar 26;4(1):57. [doi: 10.1038/s41746-021-00432-5] [Medline: 33772070]

Abbreviations

AI: artificial intelligence
HCP: health care professional
MVP: minimum viable product

Edited by A Kushniruk; submitted 13.11.23; peer-reviewed by G Carolus, L Rotteau, P Koppel; revised version received 30.08.24; accepted 14.10.24; published 11.02.25.

Please cite as:
Kalla M, O'Brien T, Metcalf O, Hoda R, Chen X, Li A, Parker C, Franco ME, Georgy S, Huckvale K, Bain C, Poon P
Understanding Experiences of Telehealth in Palliative Care: Photo Interview Study
JMIR Hum Factors 2025;12:e53913
URL: https://humanfactors.jmir.org/2025/1/e53913
doi:10.2196/53913

The Effects of Digital eHealth Versus Onsite 2-Day Group-Based Education in 255 Patients With Irritable Bowel Syndrome: Cohort Study

Birgitte Berentsen[1,2], MSc, PhD; Camilla Thuen[1], MSc; Eline Margrete Randulff Hillestad[1,2], MSc; Elisabeth Kjelsvik Steinsvik[2], MD, PhD; Trygve Hausken[1], MD, PhD; Jan Gunnar Hatlebakk[1,2], MD, PhD

[1]Department of Clinical Medicine, University of Bergen, Bergen, Norway

[2]Department of Medicine, National Center for Functional Gastrointestinal Disorders, Haukeland University Hospital, Bergen, Norway

Corresponding Author:
Birgitte Berentsen, MSc, PhD
Department of Clinical Medicine, University of Bergen, Bergen, Norway

Abstract

Background: Irritable bowel syndrome (IBS) has a high worldwide prevalence and there are few effective treatment options. Patient education can influence patient behavior that subsequently may lead to changes in attitudes and skills necessary for maintenance or improvement in management of symptom severity and quality of life. However, as postdiagnostic patient education can be resource demanding, assessment of digital approaches and verification of their effectiveness is warranted.

Objective: This cohort study aimed to investigate the effects of a digital web-based multidisciplinary eHealth program on the domains of symptom severity (Irritable Bowel Syndrome Symptom Severity Scale [IBS-SSS]), quality of life (irritable bowel syndrome quality of life [IBS-QOL]), anxiety and depression (Hospital Anxiety and Depression Scale), and a measure of general client satisfaction (client satisfaction questionnaire), compared with an onsite multidisciplinary 2-day group-based education program ("IBS-school"), in 2 cohorts of 255 patients with IBS.

Methods: Patients diagnosed with IBS, aged 15-70 years, were enrolled after referral to the Section of Gastroenterology at Haukeland University Hospital, Norway. In total, 132 patients were recruited to the eHealth program and 123 to the IBS-school group for comparison. Data were self-reported and collected digitally at enrollment and after 3 months, between 2017 and 2019. Furthermore, 71 attending the eHealth program and 49 attending the IBS-school completed the questionnaires at 3 months. Intervention response was defined as a reduction of ≥50 points on the IBS-SSS.

Results: Patients attending the eHealth program reported a significant reduction in IBS symptom severity 3 months after treatment (n=71), compared with patients attending the IBS-school (n=50). Overall, patients categorized as intervention responders in both programs showed a significant reduction in symptom severity at 3 months. Here, 41% (29/71) of patients attending the eHealth program reported a mean IBS-SSS reduction of 103 (SD 72.0) points (P<.001). In addition, these patients reported reduced anxiety (P>.001) and depression (P=.002) and enhanced quality of life (P=.03), especially the degrees of dysphoria, body image, food avoidance, health worry, interference with activity, relations, and social relations. Patients responding to the IBS-school intervention (18/50, 36%) reported a mean IBS-SSS reduction of 119 (SD 86.2) points (P<.001), and reduced depression scores (P=.046), but no difference in overall quality of life. Both groups reported the respective interventions as "good" quality health care programs, scoring them 23.5 (SD 4)—the eHealth program 23.5 (SD 4), and the IBS-school 24.2 (SD 4)—on the client satisfaction questionnaire.

Conclusions: We conclude that the digital multidisciplinary eHealth program has a significant effect on IBS symptom severity in a portion of patients; it is useful as a tool in disease self-management and does not result in worse symptom scores than an onsite multidisciplinary 2-day group-based education program after 3 months. We believe these results indicate that a digital eHealth approach is preferable to an onsite multidisciplinary 2-day group-based education program covering the same topics.

(JMIR Hum Factors 2025;12:e43618) doi:10.2196/43618

KEYWORDS

irritable bowel syndrome; IBS; eHealth; internet-guided; patient education; self-management; self-reported; patient behavior; quality of life; QOL; anxiety; depression; gastrointestinal; physiotherapist; kinesiology; cognitive behavioural therapy; CBT; Hospital Anxiety and Depression Scale; HADS; client satisfaction questionnaire; CSQ; Mann-Whitney U test; nonparametric; Wilcoxon test; neurogastroenterology

Introduction

Irritable bowel syndrome (IBS) is a chronic disorder manifested by recurrent abdominal pain and alterations in stool form or frequency [1,2]. The condition affects between 4% and 9.2% of the global population [3-5] and it is highly heterogenous. IBS' unclear etiology involves multifactorial disturbances of the bidirectional communication between the gut and the brain, including visceral hypersensitivity, low-grade inflammatory responses, intestinal motility disturbances, alterations of central nervous system processing, and alterations in gut microbiota composition [6]. However, no clear biomarker or therapeutic target for IBS has been identified. The condition lacks both a cure and medication that gives sufficient symptom relief, a fact that highlights the necessity of integrating nonpharmacological approaches including patient education in patient care [7]. Patients with IBS require personalized treatment for successful symptom relief. Approaches may include physical therapy, cognitive behavioral therapy (CBT), hypnotherapy, mindfulness and exposure therapy, and comprehensive dietary guidance by registered dietitians such as the low FODMAP (fermentable oligosaccharides, disaccharides, monosaccharides, and polyols) diet. However, access to these treatment options is often limited due to lack of trained professionals, travel distances, and cost. Thus, more accessible treatment options are warranted. Web-based interventions, patient education, and self-management has been demonstrated to be effective in patients suffering from chronic diseases, including IBS [8-10]. A systematic review from 2017 concluded with mixed results regarding the effectiveness of web-based mindfulness-based interventions compared with active control treatment conditions such as CBT. However, the study showed that treatment targeting symptoms of IBS had the largest effect size improvements [11]. A longitudinal qualitative study from 2020 showed that both telephone-based CBT and web-based CBT for IBS were positively received and had lasting positive impacts on participants' understanding of IBS symptoms, quality of life, and IBS-related behaviors [12]. A recent Japanese randomized controlled trial has shown that a multidisciplinary eHealth self-management program can reduce the severity of IBS-symptoms and improved the quality of life [13].

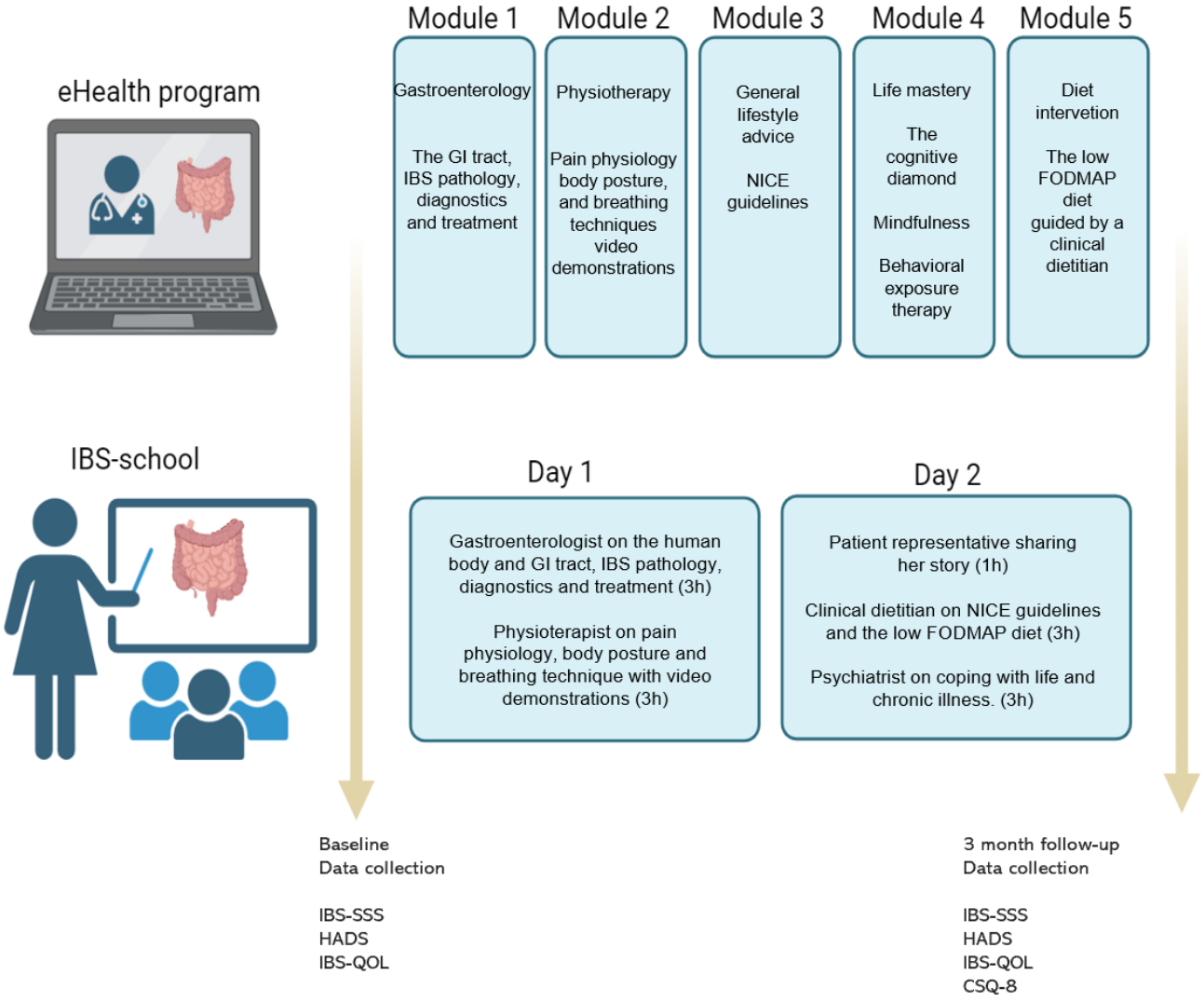
Indeed, limited health care resources, national priority guidelines, and sparse treatment options may restrict any long-term follow-up that patients with IBS may request from a secondary or tertiary care institution. Since 2012, Haukeland University Hospital has offered patients with IBS an onsite multidisciplinary 2-day group-based education program, a so-called "IBS-school." The multidisciplinary approach is designed to provide patients with evidence-based information and practical skills that may lead to improved IBS management, reduced IBS symptoms, and enhanced quality of life. Based on our clinical experience with group education, we have developed a novel internet-guided multidisciplinary self-care management program for patients with IBS, from now on referred to as the "eHealth program." The content in the eHealth program is based on topics covered in the IBS-school, but the patients also have access to digital clinical support during the program. Herein, we hypothesized that the eHealth program would have an equally good effect on symptom severity, quality of life, and patient satisfaction, compared with a cohort of patients attending IBS-school. Data from neither program have been published before. In this study, we aimed to investigate the effects of a digital web-based multidisciplinary eHealth program on the domains of symptom severity (Irritable Bowel Syndrome Symptom Severity Scale [IBS-SSS]), quality of life (irritable bowel syndrome quality of life [IBS-QOL]), anxiety and depression (Hospital Anxiety and Depression Scale [HADS]), and a measure of general client satisfaction (client satisfaction questionnaire [CSQ-8]), compared with an onsite multidisciplinary 2-day group-based education program, IBS-school, in 2 cohorts of 255 patients with IBS.

Methods

Patient Sample, Randomization, and Treatments

In this study**,** 255 patients between 15 and 70 years were recruited after being accepted for patient education at Haukeland University Hospital in Norway between 2017 and 2019. Random patients on the waiting list to the onsite multidisciplinary 2-day group-based education program, IBS-school, were contacted by phone by a study nurse and offered to attend the novel multidisciplinary digital 5-module eHealth program, from here on referred to as the "eHealth program." Patients attending the IBS-school were recruited on site when attending the 2-day group-based education program. Patients received oral and written information before giving written consent and sent it to the hospital by post. Furthermore, 123 patients that attended the IBS-school were used for comparison (N=255, 1:1). Here, patients were recruited by a study nurse on site. They received an oral and written information about the study before signing consent (Figure 1). Inclusion criteria included (1) being referred to receiving patient education on IBS by a gastroenterologist or general practitioner after the diagnosis of IBS had been determined (*International Classification of Primary Care, Second Editon* [*ICPC-2*] code D93; *International Statistical Classification of Diseases, Tenth Revision* [*ICD-10*] code K58), (2) being between 18 and 70 years, and (3) understanding written and oral Norwegian. There were no specific exclusion criteria.

Figure 1. Intervention overview. CSQ-8: client satisfaction questionnaire; FODMAP: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; GI: gastrointestinal; HADS: Hospital Anxiety and Depression Scale; IBS: irritable bowel syndrome; IBS-QOL: irritable bowel syndrome quality of life; IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale; NICE: National Institute for Health and Care Excellence.



Ethical Considerations

Eligible patients gave informed written consent and participants were given the option to withdraw from the study at any time point without a specific reason. The data were deidentified and stored on a secure hospital server, and analysis was performed on anonymous data. None of the participants were compensated. The study was approved by the Regional Ethical Committee of Western Norway (REC-2016/1098).

Interventions

The Irritable Bowel Syndrome–School

Patients attended an onsite multidisciplinary 2-day group-based education program, IBS-school, at Haukeland University Hospital. The number of participants varied between 15 and 40 each month. The program involved lectures and question and answer sessions by 4 health care professionals. Refer to Figure 1 for the intervention outline. On day 1, a gastroenterologist talked about the human body and the gastrointestinal system, what IBS is, what causes it, diagnostics, and treatment options (3 hours). Second, a physiotherapist was giving a lecture on body posture and breathing techniques including demonstrations, introducing pain physiology, and the function of the human nervous system (3 hours). On day 2, a patient representative shared her personal experience with IBS (1 hour), and a clinical dietitian gave a lecture on National Institute for Health and Care Excellence (NICE) guidelines [14] that summarizes the most recent recommendations on IBS in adults in primary care, and the low FODMAP diet [15] (3 hours). Finally, a psychiatrist gave a lecture on coping with life including health worry, social relations, tiring thoughts and feelings, adaptations, and symptoms. A nurse with specialization in gastroenterology hosted the group education program to create a comfortable environment for exchange of personal experiences and participate in group assignments.

The eHealth Program

Patients who were enrolled in the digital eHealth program had access to a comprehensive, multidisciplinary web-based program that consisted of 5 modules (refer to Figure 1 for the outline). The modules consisted of instructive texts, videos, animations, and images over 150 web-pages on a digital treatment platform by CheckWare AS [16]. In addition, patients carried out “home assignments” based on principles of CBT, a protocol by Ljótsson et al [17] at the Swedish Karolinska Institute, followed by an

optional low FODMAP diet intervention guided by a clinical dietitian. The eHealth program allowed a secure login (eg, bank-ID) and digital communication between patient and a clinical dietitian throughout the process, at the patient's own request and need. The patients were expected to finish the program over a period of 3 - 12 weeks, at his or her own pace, all dependent on their motivation and work capacity. Module contents have been described further in this study. Module 1 describes the body and the gastrointestinal system. In this first module, the patient gets an introduction to what IBS is, how it is diagnosed, and what causes it. The patient learns about the function of our digestive system and how it is regulated, and how it can be disturbed in people with IBS. Module 2 describes the posture and breathing techniques. This module focuses on the connection between IBS and musculoskeletal disorders. Many people with IBS may have a "hyperactive" nervous system that can cause pain and physical maladjustments. In this module, a physiotherapist introduces the patient to pain physiology and how the nervous system works. There is a practical section with useful exercises for people with IBS. The module will give the patient an understanding of how long-term pain occurs, why it often persists, and how to influence it. Module 3 consists of diet and lifestyle advices. There is no miracle cure for IBS, but many people experience improvement by following some general advice. In this module, we look at lifestyle advice that has been shown by research to improve symptoms in people with IBS. It is recommended to try this before eliminating other foods or following strict diets. Module 4 focuses on coping with life. In this module, the patient learns techniques from cognitive therapy including "the cognitive diamond," mindfulness, and practice systematic exposure exercises has previously shown beneficial effects for patients with IBS. The protocol has been described in a study by Ljótsson et al [17]. Module 5 describes the dietary intervention with a low FODMAP diet. This is a dietary treatment that provides symptom relief in approximately 70% of patients with IBS [18]. In this module, the patient will be introduced to the low FODMAP diet in both theory and practice. The patient had access to digital guidance by a clinical dietitian during the entire course of the study.

Questionnaires

Overview

Patients completed questionnaires related to IBS symptom severity (IBS-SSS) [2], quality of life, (IBS-QOL) [19], and anxiety and depression (HADS) [20] upon enrollment at baseline and after 3 months.

The IBS-SSS is considered the gold standard measure of IBS symptoms and contains 5 questions that measure the frequency of abdominal pain, the severity of abdominal distention, dissatisfaction with bowel habits, and interference with quality of life, scored in the range of 0 - 500. A higher score indicating worse condition, scores <175 represent mild IBS symptoms, 175 - 300 represents moderate severity, scores >300 represent severe IBS [2].

The IBS-QOL is a condition-specific measure for assessing health-related quality of life in IBS. It consists of 34 items, each with a 5-point response scale in a range of 0 - 100, where the higher score indicates a better IBS specific quality of life. There are 8 subscale scores for the IBS-QOL: dysphoria, interference with activity, body image, health worry, food avoidance, social reaction, sexual, and relationships [21].

HADS is a scale of 14 items designed to measure anxiety and depression, 7 items for each subscale (ie, anxiety and depression). The total score is the sum of the 14 items, and for each subscale the score is the sum of the respective 7 items that range from 0 to 21 [20].

CSQ-8 [22] was completed at 3 months. The questionnaire is an 8-item measure with a total score ranging from 8 to 32, with the higher number indicating greater satisfaction with treatment.

Primary Outcome Measure and Definition of Intervention Responders

In the field of IBS research, a change of 50 points in IBS-SSS score has shown to reliably indicate improvement in IBS symptom severity [2]. Our primary outcome measure was a >50-point reduction in IBS-SSS at 3 months, compared with baseline. Patients reporting a ≥50-point reduction in IBS-SSS at 3 months were categorized as "responders" to the intervention. Patients reporting <50-points on IBS-SSS were categorized as "nonresponders" to the intervention.

Statistical Analysis

Statistical analyses were performed using SPSS Statistics (version 26, IBM) for Microsoft Windows. Baseline patient characteristics and questionnaires were first presented according to intervention—eHealth group and IBS-school. At 3 months after intervention, patients were categorized as responders or nonresponders. For comparison between groups, unpaired *t* test were performed for parametric data and Mann-Whitney *U* for nonparametric data. For comparisons before and after interventions, paired *t* tests were performed for parametric data and Wilcoxon signed rank test for nonparametric data. A *P* value <.05 was considered statistically significant. The intervention responder or nonresponder analysis to the eHealth program or IBS-school was carried out performing a paired *t* test on data from patients who filled out the 3-month questionnaires (n=71 and n=50, respectively). A χ^2 test of independence was used to assess differences between intervention response.

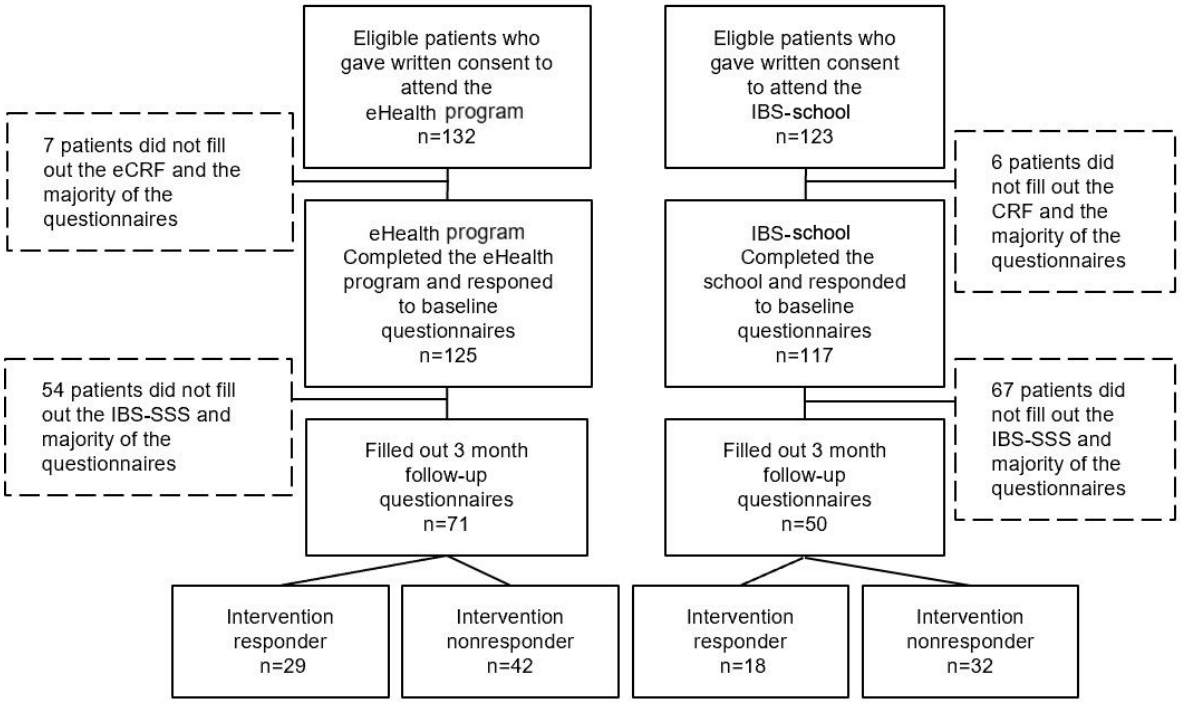
Results

Patients and Characteristics

A flowchart of participating patients is illustrated in Figure 2. Of the 132 eligible patients who gave written consent to participate in the eHealth program group of the study, 7 did not fill out the minimum requirement including the electronic case report form and the IBS-SSS questionnaire. Of the 125 patients who completed the program, 54 patients did not respond to the IBS-SSS questionnaire at 3 months. Hence, the follow-up data from patients attending the eHealth program reduced to n=71 at 3 months. Of the 123 eligible patients who gave written consent to participate in the IBS-school group of the study, 5 did not fill out the minimum requirement including the electronic case report form and the IBS-SSS questionnaire. Thus, of the 118 patients who completed the program, 69 patients did not

respond to the IBS-SSS questionnaire at 3 months. Hence, the follow-up data from patients attending the eHealth program reduced to n=49 at 3 months.

Figure 2. Flow chart of participating patients with IBS. Patients reporting a ≥50-point reduction in the IBS-SSS at 3 months were categorized as "responders to the intervention". Patients reporting <50-points on the IBS-SSS were categorized as "nonresponders to the intervention". CRF: case report form; eCRF: electronic case report form; IBS: irritable bowel syndrome; IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale.



Participants were predominantly female (190/235, 81%) with a mean age of 38.3 (SD 12.4) years (Characteristics are summarized in Table 1). An unpaired *t* test showed that there was no difference in age between groups. A Mann-Whitney *U* test was performed to evaluate whether sex differed between groups, reveling no significant difference between patients attending the eHealth program and patients attending the IBS-school (Table S1 in Multimedia Appendix 1, U=6407.0, z=−1.394, P=.16). Furthermore, participants at the IBS-school and eHealth program displayed similar baseline characteristics upon enrollment, including moderate to severe IBS symptom severity.

There was no significant difference between most relevant features at baseline, except for anxiety and depression, as shown in Table 1. Here, patients attending the eHealth program scored an average of 11 (SD 5.2) which corresponds to mild to moderate anxiety, while patients attending the IBS-school scored 9.4 (SD 5.1), corresponding to a mild level of anxiety (2-tailed t_{232}=2.37, P=.02). Thus, both patient groups displayed an anxiety score typical of a clinical case of anxiety (HADS-anxiety ≥8 is considered a clinical case [20]), and patients attending the eHealth program presented a significantly higher baseline score than patients attending the IBS-school. Furthermore, patients attending the eHealth program reported significantly higher depression scores than patients attending the IBS-school (mean 7, SD 4.1 vs mean 5.9, SD 4, respectively; t_{232}=2.01, P=.045). However, both groups displayed an average depression score corresponding to nonclinical severity of depression (HADS-depression ≤8 is considered noncase [20]) which gives the difference little relevance on group level.

Table . Baseline patient characteristics and questionnaire data of 242 patients with irritable bowel syndrome (IBS) at enrollment to the eHealth program or IBS-school interventions. Unpaired *t* test unless otherwise stated; the reduced number from total number of participants indicate missing data.

	eHealth program		IBS-school		Significance	
	n	Mean (SD)	n	Mean (SD)	*t* test (*df*)	*P* value
Participant age	125	38.3 (12.4)	117	38.4 (13.4)	−0.06 (234)	.95
Gender						
Men	27	—[a]	18	—	—	—
Women	92	—	98	—	—	.16[b]
IBS symptom severity (IBS-SSS[c])	125	282.7 (82.5)	117	268.5 (87.9)	1.29 (240)	.20
Anxiety (HADS-A[d])	124	11 (5.2)	112	9.4 (5.1)	2.37 (232)	.02[e]
Depression (HADS-D[f])	124	7 (4.1)	112	5.9 (4)	2.01 (232)	.045[e]
IBS-QOL [g]	123		117			
Overall score		50.4 (21.9)		51.8 (21.0)	−0.48 (238)	.63
Body image		41.0 (21.3)		44.6 (22.9)	−1.27 (238)	.21
Dysphoria		47.0 (23.5)		50.0 (24.0)	−0.97 (238)	.33
Food avoidance		29.0 (22.2)		28.3 (21.9)	0.25 (238)	.80
Health worry		52.5 (22.9)		57.0 (23.0)	−1.46 (238)	.15
Interference with activity		45.0 (21.9)		44.2 (21.3)	0.32 (238)	.90
Relationships		56.7 (23.2)		57.6 (22.6)	−0.27 (238)	.79
Social relations		51.6 (22.1)		54.3 (22.6)	−0.93 (238)	.35
Sexual activity		57.8 (29.6)		57.8 (32.9)	−0.05 (238)	.96

[a]Not applicable.

[b]Mann-Whitney *U* test (Mean rank, *U*, *z* in Table S1 in Multimedia Appendix 1).

[c]IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale.

[d]HADS-A: Hospital Anxiety and Depression Scale – Anxiety.

[e]Significant at the *P*<.05 level.

[f]HADS-D: Hospital Anxiety and Depression Scale – Depression.

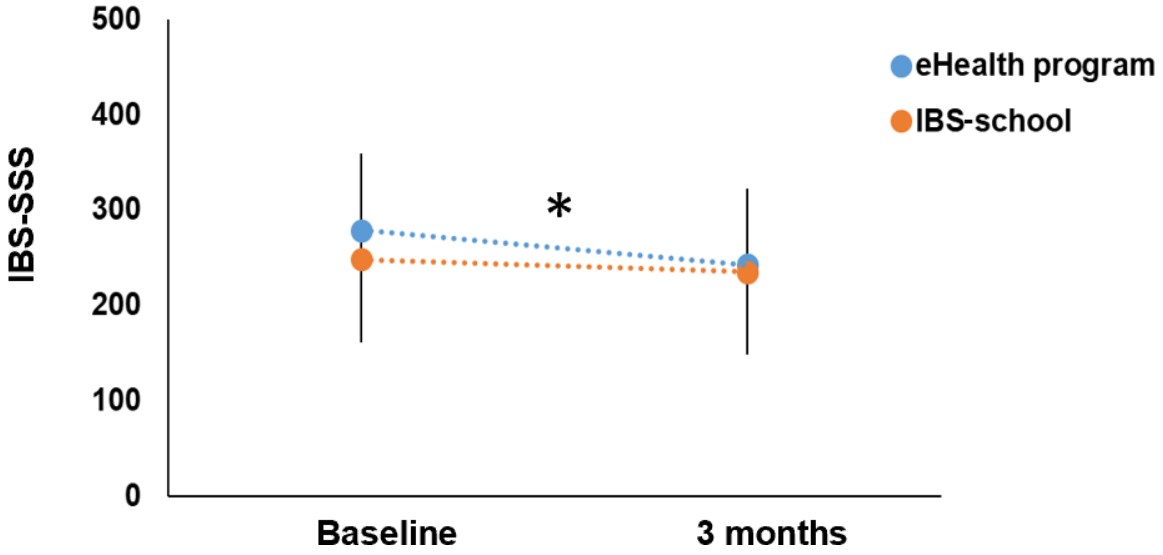
[g]IBS-QOL: irritable bowel syndrome quality of life.

Symptom Relief and Enhanced Quality of Life

Overview

A paired *t* test revealed that patients attending the eHealth program reported a significant reduction in IBS symptom severity at 3 months, shown in Figure 3 (mean 279.3, SD 80 vs mean 242.7, SD 79.3; t_{70}=4.91, *P*<.001). Comparably, patients attending the IBS-school did not report a significant reduction in symptom severity (mean 248.6, SD 86.6 vs mean 235.4, SD 98.9; t_{49}=0.85, *P*=.40). Refer to Table S2 of Multimedia Appendix 1 for data. However, none of the groups achieved a clinically meaningful improvement of ≥50 points on IBS-SSS at 3 months. Patients in the eHealth program-group reported a 37-point reduction, whereas patients attending IBS-school reported a 13-point reduction in IBS-SSS total score (Table S2 of Multimedia Appendix 1). Thus, for further analysis we categorized participants as a “responder” or “non-responder” to the intervention, where the sample response threshold for responders was set as ≥50-point decrease in total IBS-SSS score, a threshold that has been demonstrated to correlate with improvements in clinical symptoms [2].

Figure 3. Overall change in IBS-SSS over time. On average, patients who completed the 3-month questionnaires at the IBS-school reported a 13 point reduction in IBS-SSS (n= 50), whereas eHealth program participants reported a 37-point reduction (n=71). A decrease of 50 points is defined as a clinical response to the intervention, categorizing the patient as a "responder to the intervention." The IBS-SSS eHealth program score at baseline was 279 (SD 80) and 243 (SD 79) at 3 months, *P*<.001; the IBS-SSS IBS-school score at baseline was 248 (SD 87) and 235 (SD 99) at 3 months (*P*=.40). Paired *t* test results are summarized in Table S2 in Multimedia Appendix 1. **P*<.001, indicating a significant difference in IBS-SSS scores before and after eHealth program intervention. Blue dot represents the eHealth program; orange dot represents the IBS-school. Paired *t* test was performed. Error bars present SD. Lower n indicate missing data. The reduced number from the total number of participants indicate missing data. IBS: irritable bowel syndrome. IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale.



Responders

Of the 71 patients, 29 (41%) attending the eHealth program responded to the intervention with a reduction in IBS-SSS score of ≥50 points, compared with 36% (18/50) who attended the IBS-school after 3 months (Figure 3). The results in this section are summarized in Tables 2 and 3. Patients categorized as responders to the eHealth program reported a significant mean reduction in IBS-SSS score of 103 (SD 72.0) points with a moderate effect size of 0.56 (t_{28}=7.62, Cohen *d*=1.33, *P*<.001). In addition, eHealth program–responding patients reported a significant reduction in anxiety, changing from a clinical case to a noncase classification (mean −4.7, SD 6.2; t_{28}=6.89, Cohen *d*=0.85, *P*<.001). Here, the effect size of 0.39 indicated a moderate effect. On average, these patients also reported an increase in overall quality of life, summarized in Tables 2 and 3 (mean 5.80, SD 19.1; t_{28}=−2.28, Cohen *d*=−0.30, *P*=.03). Here, features such as dysphoria (mean 18, SD 23.3; t_{28}=−2.1, Cohen *d*=−0.85, *P*<.001) and body image (mean 14.8, SD 18.6; t_{28}=−3.9, Cohen *d*=−0.72, *P*<.001) increased significantly with a moderate effect size. Features such as food avoidance (mean 6.3, SD 22.5; t_{28}=−2.10, Cohen *d*=−0.13, *P*=.045), health worry (mean 10.8, SD 23.5; t_{28}=−3.44, Cohen *d*=−0.43, *P*=.002), interference with activity (mean 13.4, SD 18.8; t_{28}=−4.03, Cohen *d*=−0.63, *P*<.001), relations (mean 9.5, SD 22.9; t_{28}=−2.72, Cohen *d*=−0.37, *P*=.01), and social relations (mean 10.7, SD 19.4; t_{28}=−3.62, Cohen *d*=−0.56, *P*=.001), all improved significantly, compared with baseline. The changes in sexual activity were not significantly different after 3 months (mean 5.4, SD 29.8; t_{28}=−1.52, Cohen *d*=−0.19, *P*=.14).

Table . Changes in symptom severity, anxiety and depression, and quality of life 3 months after the eHealth program in patients with irritable bowel syndrome (IBS) who attended the eHealth program. Lower n indicate missing data; the reduced number from the total number of participants indicate missing data.

	Responder						Nonresponder					
	n	mean (SD)	*t* test (*df*)	Cohen *d*	Effect size *r*	*P* value	n	mean (SD)	*t* test (*df*)	Cohen *d*	Effect size *r*	*P* value
IBS symptom severity (IBS-SSS[a])	29	−102.8 (72.0)	7.62 (28)	1.33	0.56	<.001[b]	42	9.2 (70.4)	−1.73 (41)	−0.13	−0.06	.09
Anxiety (HADS-A[c])	29	−4.7 (6.2)	6.80 (28)	0.85	0.39	<.001[b]	39	−2.4 (4.8)	3.65 (38)	0.52	0.25	.001[b]
Depression (HADS-D[d])	29	−3.0 (4.6)	3.47 (28)	0.76	0.36	<.001[b]	39	−1.3 (3.8)	1.89 (38)	0.34	0.17	.67
IBS-QOL[e], overall score	29	5.80 (19.1)	−2.28 (28)	−0.30	−0.15	.03[b]	37	2.5 (24.8)	−0.79 (36)	−0.10	−0.05	.44
Body image		14.8 (18.6)	−3.90 (28)	−0.72	−0.34	.001[b]		3.8 (25.4)	−1.24 (36)	−0.16	−0.08	.15
Dysphoria		18.0 (23.3)	−2.10 (28)	−0.85	−0.39	<.001[b]		2.8 (24.3)	−1.20 (36)	−0.11	−0.05	.24
Food avoidance		6.3 (22.5)	−2.10 (28)	−0.26	−0.13	.045[b]		4.8 (21.7)	−1.94 (36)	−0.22	−0.11	.06
Health worry		10.8 (23.5)	−3.44 (28)	−0.43	−0.21	.002[b]		1.8 (23.5)	−0.60 (36)	−0.08	−0.04	.55
Interference with activity		13.4 (18.8)	−4.03 (28)	−0.63	−0.30	<.001[b]		2.5 (23.65)	−1.11 (36)	−0.11	−0.05	.28
Relationships		9.5 (22.9)	−2.72 (28)	−0.37	−0.18	.01[b]		4.6 (21)	−1.81 (36)	−0.20	−0.10	.08
Social relations		10.7 (19.4)	−3.62 (28)	−0.56	−0.27	.001[b]		10.7 (24.6)	−0.53 (36)	−0.46	−0.23	.60
Sexual activity		5.4 (29.8)	−1.52 (28)	−0.19	−0.10	0.14		5.4 (30.13)	−0.56 (36)	−0.19	−0.09	.58

[a]IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale.

[b]Significant at the $P<.05$ level.

[c]HADS-A: Hospital Anxiety and Depression Scale – Anxiety.

[d]HADS-D: Hospital Anxiety and Depression Scale – Depression.

[e]IBS-QOL: irritable bowel syndrome quality of life.

Table . Changes in symptom severity, anxiety and depression, and quality of life in patients with irritable bowel syndrome (IBS) 3 months after the IBS-school. Lower n indicate missing data; the reduced number from the total number of participants indicate missing data.

	Responder						Nonresponder					
	n	mean (SD)	*t* test (*df*)	Cohen *d*	Effect size *r*	*P* value	n	mean (SD)	*t* test (*df*)	Cohen *d*	Effect size *r*	*P* value
IBS symptom severity (IBS-SSS)[a]	18	−119.3 (86.2)	5.45 (17)	1.33	0.56	<.001[b]	32	46.5 (77.8)	−4.08 (31)	−0.57	−0.27	<.001[b]
Anxiety (HADS-A[c])	16	−2.5 (6.8)	1.97 (15)	0.43	0.21	.07	29	−1.9 (5.3)	2.02 (28)	0.38	0.19	.05
Depression (HADS-D[d])	16	−2.3 (4.4)	2.17 (15)	0.55	0.27	.046[b]	29	−0.5 (3.9)	0.56 (28)	0.12	0.06	.58
IBS-QOL[e], overall score	8	2.4 (27.9)	−0.49 (7)	−0.08	−0.04	.64	24	−11.4 (23.8)	3.16 (23)	5.52	0.25	.004[b]
Body image		1.68 (26.5)	−0.57 (7)	−0.05	−0.02	.59		−4.0 (22.4)	1.38 (23)	0.18	0.09	.18
Dysphoria		17 (27.2)	−6.76 (7)	−0.62	−0.30	<.001[b]		−3.3 (21.7)	1.11 (23)	0.15	0.08	.28
Food avoidance		5.3 (28)	−0.62 (7)	−0.17	−0.09	.56		0.8 (21.6)	−0.13 (23)	−0.03	−0.02	.85
Health worry		13.7 (15.3)	−2.76 (7)	−0.67	−0.32	.03[b]		0 (22.4)	−0.01 (23)	0	0	.20
Interference with activity		9.3 (28.1)	−3.14 (7)	−0.30	−0.15	.02[b]		−0.5 (19.1)	0.19 (23)	0.02	0.01	.85
Relationships		21.2 (22.6)	−3.50 (7)	−0.87	−0.40	.01[b]		18.7 (17)	−2.11 (23)	−0.86	−0.39	.046[b]
Social relations		7.8 (21.4)	−1.23 (7)	−0.27	−0.13	.26		6.4 (24.2)	1.84 (23)	0.26	0.13	.08
Sexual activity		5.4 (29.8)	1.05 (7)	−0.19	−0.10	.33		2.8 (31.6)	0.81 (23)	0.08	0.04	.43

[a]IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale.

[b]Significant at the $P<.05$ level.

[c]HADS-A: Hospital Anxiety and Depression Scale – Anxiety.

[d]HADS-D: Hospital Anxiety and Depression Scale – Depression.

[e]IBS-QOL: irritable bowel syndrome quality of life.

Patients who responded to the IBS-school intervention reported an average reduction in IBS symptom severity score of 119 (SD 86.2) points (n=18; t_{17}=5.54), Cohen d=1.33, P<.001). In addition, patients reported a reduction in depression scores (mean −2.3, SD 4.4; t_{15}=2.17, Cohen d, P=.046). However, these patients did not report significant improvements in anxiety or overall quality of life (mean 2.5, SD 6.8; t_{15}=1.97, Cohen d=0.43, P=.07 and mean 2.4, SD 27.9; t_7=−0.57, Cohen d=−0.08, P=.64, respectively). Furthermore, the number of responding patients were too low for further in-depth statistical analysis of quality of life (n=8). Data for these analyses are presented in Table S3 of Multimedia Appendix 1.

Nonresponders

Patients who were nonresponders to the eHealth program reported no improvement in symptom severity or quality of life (mean 9.2, SD 70.4; t_{41}=−1.73, Cohen d=−0.13, P=.09) and a significant decrease in anxiety scores (mean −2.4, SD 4.8, t_{38}=3.65, Cohen d=0.52, P=.001), compared with baseline. However, both changes were of small effect (r=−0.27 and 0.25, respectively). Results are summarized in Tables 2 and 3. There were no significant changes in depression scores or overall quality of life or the respective sub-categories (P>.05). Nonresponding patients attending IBS-school reported significantly enhanced symptom scores (mean 46.5, SD 77.8; t_{31}=−4.08, Cohen d=−0.57, P<.001). However, it is worth noting that these enhanced symptoms scores were not above the 50-point threshold for clinical relevance. Furthermore, these

patients reported a significantly reduced overall quality of life (mean −11.4, SD 23.8; t_{23}=3.16, Cohen *d*=5.52, *P*=.004) and a significant increase in the domain of relationships (mean 18.7, SD 17; t_{23}=−2.11, Cohen *d*=−0.86, *P*=.046). Baseline characteristics for these analyses are presented in Table S3 of Multimedia Appendix 1.

Association Between Type of Intervention and Outcome

A χ^2 test of independence analysis was used to test whether the type of intervention was independent from the intervention outcome. Results showed that the proportion of patients who responded to the intervention in both groups (29/71 vs 18/50) were the same. Hence, there was no significant association between the type of intervention and response to the intervention (Pearson χ^2_1=0.2, Φ=0.041, *P*=.65). Thus, we conclude that there is no difference in treatment response between the eHealth program and IBS-school in affecting symptom severity scores. Data from this analysis is reported in Table S4 of Multimedia Appendix 1.

Patient Satisfaction

Patient satisfaction was investigated in both groups 3 months after enrollment (CSQ-8). Patients who attended the IBS-school reported a mean score of 24.2 (SD 3.7), compared with patients attending the eHealth program with a mean score of 23.5 (SD 4.0), which are both equivalent to "good" health care offers [22]. An unpaired *t* test revealed no significant difference between group scores (t_{115}=−1.032, *P*=.31; Table 4).

Table . Difference in client satisfaction scores (CSQ-8) between patients attending the eHealth program and irritable bowel syndrome (IBS)–school. A χ^2 test of independence was conducted. The reduced number from the total number of participants indicate missing data.

		n	mean (SD)	*t* test (*df*)	Cohen *d*	Effect size r	*P* value
Overall				−1.032 (115)	−0.193	−0.01	.31
	eHealth program	68	23.46 (4.0)				
	IBS-school	49	24.20 (3.67)				

Discussion

Principal Findings

In this study, we have shown that the novel digital multidisciplinary eHealth program has a significant reducing effect on IBS symptom severity and is useful as a tool in disease self-management. In total, 41% (29/71) of participants reported significant and clinically relevant symptom relief. Furthermore, we show that the eHealth program is safe, as patients not responding to the intervention reported unchanged symptoms and quality of life at 3 months. In addition, eHealth intervention–responding patients reported significant benefits on multiple domains of IBS-related quality of life such as body image, food avoidance, health worry, interference with activity, relations, and social relations. Levels of anxiety were significantly reduced, and levels of dysphoria were improved.

Comparably, 36% (18/50) of participants reported a clinically significant effect to the onsite multidisciplinary 2-day group-based education program, the so-called "IBS-school." Thus, our results indicate that the digital multidisciplinary eHealth program may be equally effective to the IBS-school, which is often a standard treatment offer to newly diagnosed patients. Furthermore, patients responding to the IBS-school intervention did not report any significant improvements in quality of life or in anxiety, but a small not clinically meaningful decrease in depression scores. The number of IBS-school responders were too low for more in-depth statistical analysis on the domains of quality of life.

A χ^2 test of independence showed no difference between intervention outcome in the 2 groups; hence the eHealth program did not have a better intervention response than the IBS-school on the measures of symptom severity. However, the eHealth program had a significant effect on other aspects of IBS symptomatology, including anxiety and quality of life, whereas IBS-school did not.

Patients rated both the eHealth program and the IBS-school as good health care offers on measures of patient satisfaction of health care quality, scoring them 23.5 and 24.2 out of a maximum of 32 points, respectively.

We believe these results indicate that a digital eHealth approach, designed to provide patients with evidence-based information and practical skills, is preferable to an onsite multidisciplinary 2-day group-based education program covering the same topics.

Comparison With Previous Work

Initially at baseline in both groups, the greatest impairment in quality of life was observed for the subscale of food avoidance followed by body image, inactivity, and dysphoria. This order of impairment is similar to findings by Drossmann et al [21] in an international study from 2009 (n=1966). However, our findings show lower scores on all subscales except dysphoria and health worries, which are in the same magnitude. In comparison with a newer study on quality of life by Kopczyńska et al [23] (n=87), our baseline results show much lower scores on both overall and all subscales of quality of life. A recent Vietnamese study showed that patient education, lifestyle, and dietary intervention, administered by clinical pharmacists, improved IBS related quality of life compared with standard medical therapy over 8 weeks [10]. These study patients also showed a much higher baseline and postintervention quality of life-scores compared with our study. We speculate that our study participants represent a group with more severe IBS because they are referred to a tertiary health care institution that due to capacity issues has to prioritize those patients who need it the most. However, the low baseline scores on food avoidance in our study indicate that both our interventions with a key focus on diet was the right call. This core problem was targeted

because food is a known important trigger of IBS-symptoms [24]. On the domain of food avoidance, we observed an improvement in eHealth program-responders, compared with baseline (Tables 2 and 3). No such improvement was observed in patients attending the IBS-school. We speculate that this is due to the differences in comprehensiveness and durability. The eHealth program is designed to engage the patient interactively to learn how to make appropriate changes in diet. It offers a thorough guidance with videos and instructions on how to follow the low FODMAP diet, including the exclusion and reintroduction of foods. In addition, the patient has the option of digital question and answer sessions with a clinical dietitian throughout the program. The IBS-school is also designed for patients to learn how to self-help, but the program is much shorter with its 2 days of physical attendance. In the light of these results, we deduce that the eHealth program is not inferior to the established IBS-school as a health care offer. In fact, we show that an eHealth intervention program can provide the patient with self-help tools that can lead to reduced gastrointestinal symptoms and enhance quality of life for patients with IBS. This aligns with a recent randomized controlled study by Tayama et al [13] (n=40) showing that a multidisciplinary eHealth self-management program leads to an increased intake of FODMAP-groups and subsequently a more extensive diet. Severe food avoidance and dietary restriction is previously reported in 13% (829/955) of IBS-patients and this subgroup of IBS-patients reported more severe IBS-symptoms, reduced quality of life, and reduced intake of nutrients [24]. Thus, providing patients with evidence-based information, practical tools, and support by a clinical dietitian is important in clinical care of patients with IBS.

eHealth programs in the form of apps, internet-guided programs, or telehealth has recently accelerated as useful tools in clinical medicine. In an American study on satisfaction during COVID-19, most patients with IBS reported high satisfaction rates and ease of use with telehealth [25]. Here the authors reported on multiple benefits including the patient having to take less time off from work and improved access to the care team. Their most commonly reported challenges with telehealth included feeling impersonal and being unable to address all of their issues or concerns. A majority felt that telehealth was as good as or better than face-to-face visits and would use telehealth for future care. Only approximately 10% (130/1311) of the patients remained dissatisfied. In 2020, a Polish study showed that an educational program combined with elements of behavioral therapy, individualized for patients with IBS, is an important part of therapy [26]. In addition, as a part of a dietetic-led gastroenterology service in primary care, feasibility, acceptability, and cost-efficiency of using webinars to deliver first-line patient education for patients with IBS, has been shown to be successful [27]. A meta-analysis of chronic gastrointestinal illness interventions (19 studies conducted in 8 countries, n=3193) showed that eHealth gastrointestinal interventions improved patients' quality of life, psychological distress, medication adherence, and illness-related knowledge [28]. The meta-analysis also showed that eHealth gastrointestinal interventions significantly reduced the number of patient visits to the hospital. Taken together with our results, these findings support eHealth interventions holding decent promise in improving outcomes for patients with IBS.

Comparably on patient satisfaction, a randomized controlled trial by Lackner et al [29] showed that patients who received 4 gastroenterologist-led patient education sessions over 2 weeks reported 26.6 in patient satisfaction. Here, 43.5% (63/145) of patients reported improvement in addition to a higher satisfaction score than in our study. However, both these and our results reflect that subjective symptom relief may not be required for the patient to experience the treatment as useful. This is also highlighted by our patients attending the IBS-school who did not experience significant enhancement in quality of life nor reduction in symptom severity, but still reported a higher satisfaction than patients who objectively benefited more from the eHealth program.

Limitations

All 255 recruited patients completed the programs. However, there was a very high percentage that did not submit the 3-month questionnaires. In total, 41% (54/132) percent of patients attending the eHealth program, and 54% (67/123) of the patients who attended the IBS-school did not submit the 3-month follow-up questionnaires on IBS symptoms severity (Figure 1). Unfortunately, this occurred despite multiple efforts and reminders (phone calls, emails, and SMS text messages) from the research team encouraging the participants to respond. First, this high percentage of missing data are significant and may have affected the results of the study even though the rates appear similar across the groups. However, high dropout rates are a common issue in eHealth studies and known as "the law of attrition" [30]. This warrants the need for a larger study where an intention-to-treat analysis can be carried out without diminishing the power of the study. Second, eHealth literacy is defined by Norman and Skinner [31] as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem". The individual patient's level of health literacy was not mapped during the study and variations in these abilities may have affected our results. Third, the patient population represents all subtypes of IBS and there are no analyses focusing on the differences in response between patients with predominant diarrhea, constipation, or a mix of the 2. Comorbidities or other additional diagnoses are common in IBS [32] but have not been excluded in this study and may have affected the results. Fourth, the use of drugs during the study period have not been reported. Hence, many drugs have side effects such as nausea, vomiting, diarrhea, constipation, flatulence, which are symptoms that the patient may confused with IBS symptoms. Fifth, a high placebo effect, which can be up to 40%, is a known challenge in clinical studies on IBS [20]. Although this study has not been designed with a control group, the placebo effect may have affected our results in either group and have not been adjusted for. However, we may speculate that the phenomenon has affected patients in both groups equally and importantly, the placebo effect may recede after 12 weeks, which was our end point [33]. Sixth, as both programs are broad and cover a variety of information, advice, and treatment including the low FODMAP diet and principles of CBT, another limitation of this study is the unknown specifics patients were

responding to in either program. Indeed, there are no measures on cognition verifying the direct effects of the principles of CBT or exposure therapy. This needs to be further investigated in a prospective study, and we acknowledge that an in-depth explanation for our observed benefits after attending the eHealth program remain to be clarified. Thus, these aspects are objectives in our currently ongoing randomized controlled trial [34]. In the light of limited primary and secondary health care resources, it will be useful to develop prediction tools to identify which patients may achieve improvement in both symptom severity and domains of quality of life. For these stratification analyses to be clinically meaningful, the number of participants need to be higher than reported in this study and performed in and randomized controlled trial.

Conclusions

We conclude that the digital multidisciplinary eHealth program has a significant effect on IBS symptom severity in a portion of patients, and is useful as a tool in disease self-management. In addition, it does not result in worse symptom scores than an onsite multidisciplinary 2-day group-based education program after 3 months. We believe these results indicate that a digital eHealth approach, that include benefits such as 3 months unlimited access to quality assured information and treatment with documented effect, is preferable to an onsite multidisciplinary 2-day group-based education program covering the same topics.

Acknowledgments

The authors wish to thank all patients who participated at the irritable bowel syndrom (IBS)–school and the eHealth program. In addition, we wish to thank Haukeland University Hospital's “Learning and Mastering Center” and their wonderful employees, the patient educators psychiatrist Tor Jacob Moe, physiotherapist Eirik Østhammer, psychosomatic psychiatrist Jørn Bødtker, patient representative Cecilie Ågotnes, IBS-school host nurse Per Førde Refsnes, clinical dietitians Cecilie Larsen and CT, CheckWare AS, and eHealth program developer Jan Gunnar Hesthammer, Section for eHealth, Haukeland University Hospital. We also wish to thank the Department of Medicine and Section for Gastroenterology at Haukeland University Hospital, Tine Nordgreen and Research Centre for Digital Mental Health Services, Centre for Research-Based Innovation, Haukeland University Hospital, and Bjrann Ljottson, Karolinska Institutet, and Pelle Lindfors, Sahlgrenska, Sweden. This work was supported by Helse Vest under grants HV1148 and R-10923.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables.
[XLSX File, 47 KB - humanfactors_v12i1e43618_app1.xlsx]

References

1. Ford AC, Sperber AD, Corsetti M, Camilleri M. Irritable bowel syndrome. Lancet 2020 Nov 21;396(10263):1675-1688. [doi: 10.1016/S0140-6736(20)31548-8] [Medline: 33049223]
2. Francis CY, Morris J, Whorwell PJ. The irritable bowel severity scoring system: a simple method of monitoring irritable bowel syndrome and its progress. Aliment Pharmacol Ther 1997 Apr;11(2):395-402. [doi: 10.1046/j.1365-2036.1997.142318000.x] [Medline: 9146781]
3. Cassar GE, Youssef GJ, Knowles S, Moulding R, Austin DW. Health-related quality of life in irritable bowel syndrome: a systematic review and meta-analysis. Gastroenterol Nurs 2020;43(3):E102-E122. [doi: 10.1097/SGA.0000000000000530] [Medline: 32487960]
4. Oka P, Parr H, Barberio B, Black CJ, Savarino EV, Ford AC. Global prevalence of irritable bowel syndrome according to Rome III or IV criteria: a systematic review and meta-analysis. Lancet Gastroenterol Hepatol 2020 Oct;5(10):908-917. [doi: 10.1016/S2468-1253(20)30217-X] [Medline: 32702295]
5. Sperber AD, Bangdiwala SI, Drossman DA, et al. Worldwide prevalence and burden of functional gastrointestinal disorders, results of Rome Foundation Global Study. Gastroenterology 2021 Jan;160(1):99-114. [doi: 10.1053/j.gastro.2020.04.014] [Medline: 32294476]
6. Hillestad EMR, van der Meeren A, Nagaraja BH, et al. Gut bless you: the microbiota-gut-brain axis in irritable bowel syndrome. World J Gastroenterol 2022 Jan 28;28(4):412-431. [doi: 10.3748/wjg.v28.i4.412] [Medline: 35125827]
7. Camilleri M. Diagnosis and treatment of irritable bowel syndrome: a review. JAMA 2021 Mar 2;325(9):865-877. [doi: 10.1001/jama.2020.22532] [Medline: 33651094]
8. Ankersen DV, Weimers P, Bennedsen M, et al. Long-term effects of a web-based low-FODMAP diet versus probiotic treatment for irritable bowel syndrome, including shotgun analyses of microbiota: randomized, double-crossover clinical trial. J Med Internet Res 2021 Dec 14;23(12):e30291. [doi: 10.2196/30291] [Medline: 34904950]

9. Lindfors P, Axelsson E, Engstrand K, et al. Online education is non-inferior to group education for irritable bowel syndrome: a randomized trial and patient preference trial. Clin Gastroenterol Hepatol 2021 Apr;19(4):743-751. [doi: 10.1016/j.cgh.2020.04.005] [Medline: 32289541]
10. Thong VD, Ngoc Phuc N, Quynh BTH. Effectiveness of educational intervention carried out by clinical pharmacists for the quality of life of patients with irritable bowel syndrome: a randomized controlled trial. JGH Open 2021 Feb;5(2):242-248. [doi: 10.1002/jgh3.12477] [Medline: 33553662]
11. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303. [doi: 10.2196/jmir.7487] [Medline: 28860106]
12. Hughes S, Sibelli A, Everitt HA, et al. Patients' experiences of telephone-based and web-based cognitive behavioral therapy for irritable bowel syndrome: longitudinal qualitative study. J Med Internet Res 2020 Nov 20;22(11):e18691. [doi: 10.2196/18691] [Medline: 33216002]
13. Tayama J, Hamaguchi T, Koizumi K, et al. Efficacy of an eHealth self-management program in reducing irritable bowel syndrome symptom severity: a randomized controlled trial. Sci Rep 2024 Jan 3;14(1):4. [doi: 10.1038/s41598-023-50293-z] [Medline: 38172498]
14. Hookway C, Buckner S, Crosland P, Longson D. Irritable bowel syndrome in adults in primary care: summary of updated NICE guidance. BMJ 2015 Feb 25;350:h701. [doi: 10.1136/bmj.h701] [Medline: 25716701]
15. Ireton-Jones C. The low FODMAP diet: fundamental therapy in the management of irritable bowel syndrome. Curr Opin Clin Nutr Metab Care 2017 Sep;20(5):414-419. [doi: 10.1097/MCO.0000000000000398] [Medline: 28617710]
16. CheckWare. URL: https://www.checkware.com [accessed 2025-01-14]
17. Ljótsson B, Hesser H, Andersson E, et al. Provoking symptoms to relieve symptoms: a randomized controlled dismantling study of exposure therapy in irritable bowel syndrome. Behav Res Ther 2014 Apr;55:27-39. [doi: 10.1016/j.brat.2014.01.007] [Medline: 24584055]
18. Whelan K, Martin LD, Staudacher HM, Lomer MCE. The low FODMAP diet in the management of irritable bowel syndrome: an evidence-based review of FODMAP restriction, reintroduction and personalisation in clinical practice. J Hum Nutr Diet 2018 Apr;31(2):239-255. [doi: 10.1111/jhn.12530] [Medline: 29336079]
19. Patrick DL, Drossman DA, Frederick IO, DiCesare J, Puder KL. Quality of life in persons with irritable bowel syndrome: development and validation of a new measure. Dig Dis Sci 1998 Feb;43(2):400-411. [doi: 10.1023/a:1018831127942] [Medline: 9512138]
20. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
21. Drossman DA, Morris CB, Schneck S, et al. International survey of patients with IBS: symptom features and their severity, health status, treatments, and risk taking to achieve clinical benefit. J Clin Gastroenterol 2009 Jul;43(6):541-550. [doi: 10.1097/MCG.0b013e318189a7f9] [Medline: 19384249]
22. Larsen DL, Attkisson CC, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
23. Kopczyńska M, Mokros Ł, Pietras T, Małecka-Panas E. Quality of life and depression in patients with irritable bowel syndrome. Prz Gastroenterol 2018;13(2):102-108. [doi: 10.5114/pg.2018.75819] [Medline: 30002768]
24. Melchior C, Algera J, Colomier E, Törnblom H, Simrén M, Störsrud S. Food avoidance and restriction in irritable bowel syndrome: relevance for symptoms, quality of life and nutrient intake. Clin Gastroenterol Hepatol 2022 Jun;20(6):1290-1298. [doi: 10.1016/j.cgh.2021.07.004] [Medline: 34229035]
25. Yu C, Tovar S, Shi J, et al. Satisfaction with telehealth among patients with irritable bowel syndrome during COVID-19: cross-sectional survey findings. Clin Transl Gastroenterol 2022 Sep 1;13(9):e00515. [doi: 10.14309/ctg.0000000000000515] [Medline: 35981244]
26. Sierżantowicz R, Lewko J, Jurkowska G. The impact of an individual educational program on the quality of life and severity of symptoms of patients with irritable bowel syndrome. Int J Environ Res Public Health 2020 Jun 13;17(12):4230. [doi: 10.3390/ijerph17124230] [Medline: 32545800]
27. Williams M, Barclay Y, Harper L, Marchant C, Seamark L, Hickson M. Feasibility, acceptability and cost efficiency of using webinars to deliver first-line patient education for people with irritable bowel syndrome as part of a dietetic-led gastroenterology service in primary care. J Hum Nutr Diet 2020 Dec;33(6):758-766. [doi: 10.1111/jhn.12799] [Medline: 32816367]
28. Rohde JA, Barker JO, Noar SM. Impact of eHealth technologies on patient outcomes: a meta-analysis of chronic gastrointestinal illness interventions. Transl Behav Med 2021 Feb 11;11(1):1-10. [doi: 10.1093/tbm/ibz166] [Medline: 31731292]
29. Lackner JM, Jaccard J, Keefer L, et al. Improvement in gastrointestinal symptoms after cognitive behavior therapy for refractory irritable bowel syndrome. Gastroenterology 2018 Jul;155(1):47-57. [doi: 10.1053/j.gastro.2018.03.063] [Medline: 29702118]
30. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

31. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
32. Shiha MG, Aziz I. Review article: physical and psychological comorbidities associated with irritable bowel syndrome. Aliment Pharmacol Ther 2021 Dec;54 Suppl 1:S12-S23. [doi: 10.1111/apt.16589] [Medline: 34927759]
33. Spiller RC. Problems and challenges in the design of irritable bowel syndrome clinical trials: experience from published trials. Am J Med 1999 Nov 8;107(5A):91S-97S. [doi: 10.1016/s0002-9343(99)00086-8] [Medline: 10588179]
34. Digital treatment of irritable bowel syndrome (IBS) (DIGIBS). National Library of Medicine. 2023. URL: https://clinicaltrials.gov/study/NCT06117865?cond=Irritable%20Bowel%20Syndrome&term=digibs&rank=1 [accessed 2025-01-14]

Abbreviations

CBT: cognitive behavioral therapy
CSQ-8: client satisfaction questionnaire
FODMAP: fermentable oligosaccharides, disaccharides, monosaccharides, and polyols
HADS: Hospital Anxiety and Depression Score
IBS: irritable bowel syndrome
IBS-QOL: irritable bowel syndrome quality of life
IBS-SSS: Irritable Bowel Syndrome Symptom Severity Scale
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
***ICPC-2*:** *International Classification of Primary Care, Second Edition*

Edited by A Kushniruk; submitted 26.10.22; peer-reviewed by A Al-Asadi, I Mesters, S Hughes; revised version received 31.08.24; accepted 14.10.24; published 03.02.25.

Please cite as:
Berentsen B, Thuen C, Hillestad EMR, Steinsvik EK, Hausken T, Hatlebakk JG
The Effects of Digital eHealth Versus Onsite 2-Day Group-Based Education in 255 Patients With Irritable Bowel Syndrome: Cohort Study
JMIR Hum Factors 2025;12:e43618
URL: https://humanfactors.jmir.org/2025/1/e43618
doi:10.2196/43618

Original Paper

Unveiling Sociocultural Barriers to Breast Cancer Awareness Among the South Asian Population: Case Study of Bangladesh and West Bengal, India

Fahmida Hamid[1*], PhD; Tania Roy[1*], PhD

New College of Florida, Sarasota, FL, United States

[*]all authors contributed equally

Corresponding Author:
Fahmida Hamid, PhD
New College of Florida
5800 Bay Shore Rd
Sarasota, FL, 34243
United States
Phone: 1 (941) 487 5000
Fax: 1 (941) 487 5000
Email: fhamid@ncf.edu

Abstract

Background: Bangladesh and West Bengal, India, are 2 densely populated South Asian neighboring regions with many socioeconomic and cultural similarities. In dealing with breast cancer (BC)–related issues, statistics show that people from these regions are having similar problems and fates. According to the Global Cancer Statistics 2020 and 2012 reports, for BC (particularly female BC), the age-standardized incidence rate is approximately 22 to 25 per 100,000 people, and the age-standardized mortality rate is approximately 11 to 13 per 100,000 for these areas. In Bangladesh, approximately 90% of patients are at stages III or IV, compared with 60% in India. For the broader South Asian population, this figure is 16%, while it is 11% in the United States and the United Kingdom. These statistics highlight the need for an urgent investigation into the reasons behind these regions' late diagnoses and treatment.

Objective: Early detection is essential for managing BC and reducing its impact on individuals. However, raising awareness in diverse societies is challenging due to differing cultural norms and socioeconomic conditions. We aimed to interview residents to identify barriers to BC awareness in specific regions.

Methods: We conducted semistructured interviews with 17 participants from West Bengal and Bangladesh through Zoom (Zoom Video Communications). These were later transcribed and translated into English for qualitative data analysis. All our participants were older than 18 years, primarily identified as female, and most were married.

Results: We have identified 20 significant barriers to effective BC care across 5 levels—individual, family, local society, health care system, and country or region. Key obstacles include neglect of early symptoms, reluctance to communicate, societal stigma, financial fears, uncertainty about treatment costs, inadequate mental health support, and lack of comprehensive health insurance. To address these issues, we recommend context-specific solutions such as integrating BC education into middle and high-school curricula, providing updates through media channels like talk shows and podcasts, promoting family health budgeting, enhancing communication at cultural events and religious gatherings, offering installment payment plans from health care providers, encouraging regular self-examination, and organizing statewide awareness campaigns. In addition, social media can be a powerful tool for raising mass awareness while respecting cultural and socioeconomic norms.

Conclusions: Fighting BC or any fatal disease is challenging and requires support from various dimensions. However, studies show that raising mass awareness is crucial for the early detection of BC. By adopting a sensitive and well-informed approach, we aim to improve the early detection of BC and help reduce its impact on South Asian communities.

(JMIR Hum Factors 2025;12:e53969) doi:10.2196/53969

KEYWORDS

Bangladesh; West Bengal; India; Asia; breast cancer; awareness; early detection; screening; sociocultural barriers; health knowledge; cultural; stigma; social media; socioeconomic

Introduction

According to the GLOBOCAN (Global Cancer Observatory) 2020 report on the estimates of cancer incidence and mortality among the people of 185 countries (produced by the International Agency for Research on Cancer), female breast cancer (BC) has been identified as the most commonly diagnosed cancer (11.7%) and is the fifth leading cause of cancer mortality worldwide [1]. The same report suggests that among women, BC accounts for 1 in 4 cancer cases and for 1 in 6 cancer deaths, ranking first for incidence in most of the countries (159 of 185 countries). Through numerous studies, researchers found that early detection can significantly reduce fatal outcomes, and awareness is one of the keys to early detection [2]. Spreading awareness is a challenge for different societies for various reasons, and researchers have been working on identifying those for a long time. In this work, we assess the barriers to mass awareness (and thus far early detection) of BC among people in 2 very densely populated but culturally similar and geographically neighboring areas—Bangladesh and West Bengal, India.

Bangladesh (area: 57,321 square miles) and West Bengal, India (area: 34,267 square miles) are 2 territories in South Asia. Bangladesh's official language is Bengali, and the mass people primarily use the Bengali language for daily conversation. In total, 8% (91,276,115/1,210,569,573) of the Indian population speaks Bengali and mainly resides east of the country, with a concentration in West Bengal [3]. The population density of Bangladesh is 1328.68 people per sq km (collected from the web), and in West Bengal, India is 1028 people per sq km (according to their 2021 census). The deep sociocultural linkages between the Bengali linguistic community in eastern India and Bangladesh remain stable even after India's partition and independence in 1947. The unemployment rate in Bangladesh is 5.23%, and in West Bengal, it is 5.2% (as of June 2022). A total of 70.54% (64,385,546/91,276,115) of people from West Bengal follow Hinduism, and 27.01% (24,654,825/91,276,115) follow Islam. In Bangladesh, 91.04% (150,360,405/169,828,921) are Muslims, and 7.95% (13,130,109/169,828,921) are Hindus. The life expectancy in West Bengal, India, is 71.2 years [4], and in Bangladesh, it is 73.57 years [5]. The approximate sex ratio (as of 2022) in Bangladesh is 102.12 males per 100 females, whereas in West Bengal, it is 105 males for every 100 females. The average literacy in Bangladesh is 74.91% [6], and in West Bengal is 76.26% [6]. The commonality in the history and sociocultural environment of eastern India and Bangladesh has created a pseudo-homogeneous society where language is one of the fundamental elements in this connectivity. Hence, our work focuses on understanding attitudes, perceptions, and behaviors around BC-related issues in these 2 densely populated areas.

Unfortunately, there are no available sources of comprehensive epidemiological, pathological, and outcomes data for patients with BC [7,8] for the chosen regions; so, we are using the closest information from the GLOBOCAN 2020 report [1] to estimate the incidence and mortality rate of BC in this part of the world (Table 1). The factors contributing to the rising incidence and mortality of BC in low-income countries such as Bangladesh and West Bengal, India, are multifaceted. It is important to acknowledge that the solutions required for addressing these challenges differ significantly from those implemented in high-income countries [9]. Therefore, adopting models or policies directly from high- or middle-income countries, which have demonstrated success in tackling BC, may not be applicable in this context.

Table 1. Age-standardized incidence rate and age-standardized mortality rate of breast cancer in India and Bangladesh (GLOBOCAN 2020 report) and some relevant sources like the World Health Organization.

Scope	Age-standardized incidence rate in 100,000	Age-standardized mortality rate in 100,000
World	47.8 [1]	13.6 [1]
South Central Asia (Bangladesh and India are geographically part of the area)	26.2 [1]	13.1 [1]
West Bengal, India	25.2 [8]	13.62 [10]
Bangladesh	21.4 [8]	10.69 [11]

Despite being classified as least developed countries, Bangladesh and West Bengal, India, are home to several highly equipped and specialized cancer hospitals, some for BC only; for example, the National Institute of Cancer Research and Hospital (Dhaka, Bangladesh) or Tata Medical Center (Kolkata, West Bengal, India). These institutions, staffed with highly skilled and experienced surgeons, oncologists, and gynecologists, offer mammograms, CT (computed tomography) scans, and other modern screening and medical facilities. However, advanced medical facilities are scarce in rural and suburban areas, and the capacity of existing institutions often falls short of meeting the demand. In addition, there are also significant uncertainties regarding the overall cost of treatment, the duration of care, mental health support, and gaps in preventative measures. Given this context, our goal is to identify key gaps in existing approaches to raising awareness about BC and addressing the barriers that prevent people from seeking preventative care and timely treatment. By targeting these blind spots, we aim to make meaningful changes that promote early detection and improve access to necessary medical services.

In this work, to understand the barriers to BC awareness, we conducted an open-ended interview with the residents from both regions about BC awareness (to assess their current state of knowledge, attitude and readiness, awareness, and so on) After carefully analyzing our interview data, we highlighted some common obstacles in both regions or societies, by grouping

them into different societal levels such as individual, family, local society, health care system, and state or country. In the upcoming sections, we discuss the study design (Methods), the highlighted barriers with some root causes, and suggested solutions (Results); then, we suggest how technology and social media can play a crucial role in resolving some of those effectively (Discussion).

Methods

Study Design

The main objective was to evaluate the general knowledge and awareness of the participants regarding BC, identify sociocultural factors that could hinder mass awareness campaigns, and examine health care disparities and systemic issues. We conducted individual interviews with the participants using a semi structured questionnaire that included a variety of specific questions as well as 1 open-ended question (Table 2). The participants' responses included personal experiences and anecdotal stories.

Table 2. Our questionnaire.

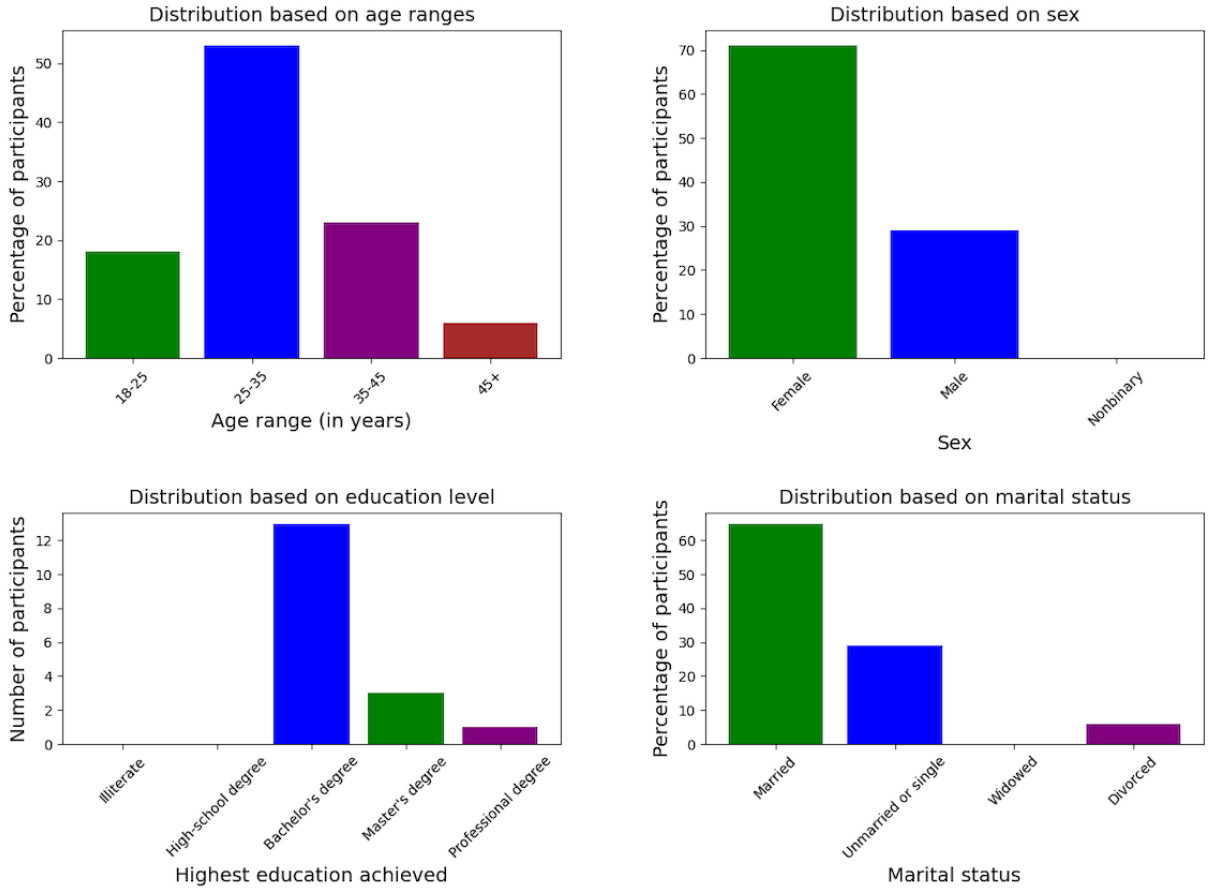
Category	Questions
Awareness	• People of which age range do you think are at high risk of getting breast cancer? • What are the most common symptoms of Breast Cancer? • What are the common misconceptions about Breast Cancer? • Are there any media-related programs that have provided information regarding Breast Cancer or Breast Cancer diagnosis or progression of the disease? Example: Television talk shows, YouTube, WhatsApp, Facebook groups, or movie or television shows • Are you aware that breast cancer can affect anybody regardless of their sex and gender classification? If not, what gender and sex demographic do you think have the possibility of getting Breast Cancer?
Self-diagnosis	• Do you know and voluntarily do the Breast test yourself at home?
Post-diagnosis support	• What were the support mechanisms (nonpharma) provided to you? For example, mental health or NGO support, nutrition, or self-care. • Was the financial impact of the proposed treatment plan discussed with the doctor?
Open-ended	• Is there anything you would like to share about your journey or experience with Breast Cancer?

[a]NGO: nongovernmental organization.

Participant Selection and Recruitment

We recruited participants through snowball sampling [12], using mutual connections and social media platforms such as Facebook (Meta). Semistructured interviews were conducted with 17 participants from West Bengal and Bangladesh. All our participants were above the age of 18 years, and recruitment was focused on participants who resided (at the time of study) in the 2 regions mentioned above. Our participants primarily identified as female, with an average age of 25-35 years, and most were married (Figure 1). The participants selected the language during the interview; most spoke Bengali. The researchers conducted the study through Zoom (Zoom Video Communications), and the audio recordings were transcribed and translated into English for data analysis.

Figure 1. Distribution of participants.



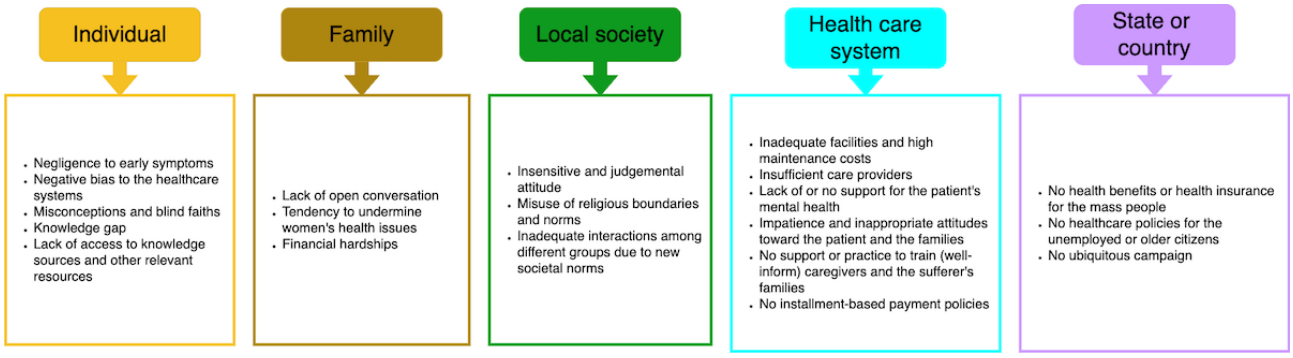
Demographics

We aimed to recruit a representative sample of participants from diverse educational and financial backgrounds from the region, as socioeconomic status and education play a huge role in the participants' access to health care in the region. Also, a cultural stereotype related to educational qualifications is widely propagated, relating educational accomplishments to open-mindedness, access to health care, and overall "success" in life [13]. Our participants self-declared their financial status as middle- and high-economic classes and were all in professions such as business, architecture, social development work, or education. Most of our participants had completed their undergraduate degrees (n=13), few had graduate degrees (n=3), and some were currently working toward a professional degree (n=1). The socioeconomic status of our participants is important in the larger context of interpreting the results and will be discussed further in the upcoming section.

Data Collection and Analysis

We transcribed the interview audio recordings, which were between 15 and 45 minutes long, and translated into English. Our qualitative data analysis follows a tiered approach and an inductive methodology [14,15] to generate themes. In tier 1, the initial coder examined the transcript identifying recurring themes, clustering them based on similarity, and using the descriptive coding methodology to label them. This clustering formed the foundation for subsequent analysis. Moving to tier 2, a second coder independently reviewed the clustered themes and descriptive terms associated with them, providing validation by either corroborating existing clusters or proposing new ones as necessary. Finally, tier 3 involved collaborative discussions between both coders to assess the relevance and significance of the clustered themes and descriptive terms associated with them. Through iterative deliberations and consensus-building, the final classification of themes was determined (Figure 2). As our interviews were semistructured (where the participants had the freedom to take the narrative and provide more information beyond a fixed set of questions), this method of generating themes bottom up and assigning descriptive terms to it allowed us the flexibility to show nuance and highlight subthemes. Both coders were familiar with the language spoken by the participants (Bengali) to parse through the linguistic, cultural, and social significance of the data generated during the interview sessions.

Figure 2. Outcome of our qualitative data analysis: major barriers to breast cancer awareness in the region.



The hierarchical representation of our themes and subsequent categories (Figure 2) is inspired by the presentation of the barrier tree [16] developed for a mobile-based BC monitoring study for the people in rural Bangladesh around 2012. Notably, several factors (such as feeling shy, fear, lack of familiarity, undermining women problems, religious belief, shared beliefs and practices, communication problems, scarcity of doctors, inconsistent patient data, long-term monitoring, and poverty) identified then are still highly relevant and came up during our analysis. The distinctions between both works are listed below.

First, most of our participants are city dwellers, but they shared stories and experiences from different parts (city, suburb, or rural) of the country or state during their interviews. Haque et al [16] studied in rural setups about 1.2 decades ago.

Second, we present the barriers by clustering them into 5 socioeconomic themes (individual, family, local society, health care system, and country or state). Haque et al [16] presented their findings into categories like identification and disclosure, achieving treatment, continuation of treatment, environmental issues, and user issues.

Third, while many research articles in the domain of BC primarily adopt a medical science perspective, such as the study by Haque et al [16], where the team visited a clinic or hospital to gather data and generate a barrier tree, our approach diverges by examining BC through a societal framework. By doing so, we aim to offer a comprehensive analysis that encompasses both the medical and societal impacts of BC. This approach allows us to explore the same problem from different lenses, shedding light on the multifaceted challenges and implications of BC beyond the purely medical aspect.

Finally, as a novel addition to the field, we delve into the mental health relevancy of patients with cancer and their caregivers, a crucial aspect that was not addressed in the other work.

We discuss the outcome of the qualitative analysis in the next section.

Ethical Considerations

The institutional review board of New College of Florida approved the study (Protocol #22-037), and we strictly followed all protocols to ensure data privacy and integrity, demonstrating our commitment to ethical standards. For the study, participation was voluntary; participants did not have to participate and could stop at any time. There were no penalties or loss of benefits or opportunities if the participant did not participate or decided to stop once started. Their decision to participate or not to participate did not affect their job status, employment record, employee evaluations, or advancement opportunities. The participants did not receive any benefit from the research team. There was no cost to participate. This research was considered minimal risk. Minimal risk means that the study risks are the same as those in daily life.

Results

The results of the data analysis are discussed in 2 phases; first, barrier identification and second, awareness raising.

Barrier Identification

The majority of our participants were aware of BC's symptoms, possible treatment mechanisms, costs, survival rate, and so on. During the interviews, they were able to list most of the significant symptoms of BC, like deformed breasts or nipples, lumps in the breast and armpit, discharge of pus or liquids, irritation, flaky breast skin, irregular breast pains, overall fatigue and weight loss, and so on. All but 3 participants knew of at least 1 person who has or had BC—either in their own or extended family, in the neighborhood, or in a friend's family—indicating how predominant this disease has become (Tables 3 and 4). A total of 14 participants "had heard of" self–breast-examinations, 13 participants "knew how to" perform self–breast-examinations, and 7 participants "had performed" self–breast-examinations (Figure 3). A greater than 50% decrease in the number of participants who had conducted self–breast examination is noteworthy.

Table 3. Number of responses to the question: "How many patients with breast cancer do you know of?"

Percentage of participants	Number of patients or survivors with BC[a]
18	0
35	1
35	2
6	3
6	4

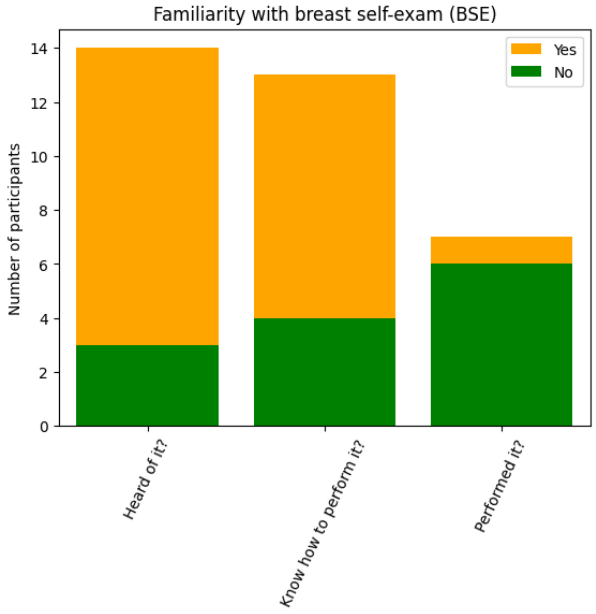
[a]BC: breast cancer.

Table 4. Patients with breast cancer in immediate family.

Percentage of participants	Number of patients or survivors with BC[a] in immediate family
59	0
24	1
17	2

[a]BC: breast cancer.

Figure 3. Familiarity test for "breast self-examination.".



The following sections share the barriers we highlighted from our data. Since the barriers often have cause-and-effect relationships and circular causality, we will not present them according to the identified societal levels like in Figure 2; instead, we would discuss them under the same section or title if they mostly appeared together during the interviews.

Tendency to Neglect Early Symptoms and Knowledge Gap

Our survey participants agreed that although BC is well-known, many people are unaware of the symptoms and often miss early signs. Specifically, when trying to identify symptoms in themselves, they either don't know what to look for or tend to disregard early signs of discomfort. For example, 1 participant stated:

> *One day, my mom had severe stomach pain. We took her to the doctor, who asked us to complete a USG of her abdomen as the doctor suspected that my mom had a tumor in her stomach. While doing the USG for the tumor, the doctor detected that she had gotten breast cancer, not a tumor…… it was already a late detection. It's at a deadly stage already. The treatment would have been much less painful if she had been diagnosed earlier.* [b007]

A similar story also stated the reluctance to pay heed to early symptoms or disclose them:

> *Though she (a neighbor) suspected it, she did not disclose it initially. When the pain became unbearable, then only she went to the doctor. She went to the doctors in the middle stage (not the early stage).* [b011]

> *One of my aunts (a family friend) had been diagnosed with breast cancer. However, it had already spread to the throat and jaw when she was diagnosed. So, the doctors could not remove/cure it, and she passed away.* [b008]

These responses reflect the results Amin et al [17] found in a survey that the research team conducted in Bangladesh in 2020; out of the 9 symptoms of BC, approximately 50% (N=500) of participants were unaware of 5 or more symptoms.

Misconceptions

Although none of our participants were medical professionals, our participant pool consisted of individuals who had all completed at least an undergraduate degree, were socioeconomically affluent, and were self-professed, open-minded individuals. Due to their educational backgrounds, we assumed that the number of misconceptions about BC would be lower than other groups. However, our participants had several deep-rooted misconceptions about BC, which are listed below.

First, it can only afflict females. Surprisingly, 10 out of 17 participants (59%) believed BC could only afflict females. The rest of the participants agreed BC occurred, irrespective of sex.

Second, it is more likely to afflict married women and those who have recently become mothers. Several participants stated that in married women, especially after giving birth to a child or during the breastfeeding period, it is highly likely to get inflicted with BC.

Third, it is more likely to afflict older women. Several participants responded that BC only affects older women; when probed further, they believed that the age range between 25 and 45 was not particularly susceptible to BC. Reports have suggested that Indian women in their early thirties till fifties are at considerable risk of developing BC, and the incidence risk increases till its peak when they reach 50-64 years of age. Furthermore, 1 in 28 Indian women are likely to develop BC during their lifetime [18-21]. It is more (1 in 22) for urban women than for the rural group (1 in 60). Amin et al [17] report similar statistics for Bangladesh (22.5 per 100,000 women and mean age being 41.8 years and 56% of the diagnoses were for reproductive-age women).

At the end of our interviews, we informed our participants that (1) BC is statistically more dominant among females but can afflict anyone, irrespective of gender and (2) there are various types (metastatic, inflammatory, and so on) of BCs that afflict people from various age ranges, so directly linking it to a particular age, marital status, or maternity may not be the best idea.

These factual misconceptions and knowledge gaps about BC are major factors related to the diagnosis of BC at an early stage. Our participant responses reinforced the need to disseminate accurate information to raise awareness.

Social Stigma and Lack of Open Conversation

First, beauty and breasts—judgmental attitudes. In addition to being an organ, breasts play a significant role in defining femininity and women's role in South Asian cultures. The fear of the disease attacks the fundamentals of female identity as a woman's breast conjures her sexuality and her capacity to nurture [22]. In patriarchal societies, such as Bengalis in West Bengal and Bangladesh, irrespective of educational or financial gains and social liberties through the years, women are perceived to have the roles of mothers, wives, or homemakers, whereas men are mostly seen as primary breadwinners of the family [23].

According to a report by a Pakistani Non-Governmental Organization, Aurat Foundation and USAID (United States Agency for International Development; 2016), in South Asian cultures, feminine traits include emotional, nurturing, passive, good homemakers, fair-skinned, without body or facial hair, long-haired, with a narrow waist, and sexually submissive. In contrast, masculine traits include being assertive, primary breadwinners, risk-taking, rational, and brave [24]. The ultimate danger of this notion to a Bengali (or maybe many people from other parts of the world) is that they feel one becomes "lesser feminine" if one has any breast health-related issue. On top of that, many people assume that BC's ultimate aftereffect is mastectomy, if not death, which adds a new level of stigma [25,26].

> *When I was very young, I learned that one of my friend's mom had breast cancer. At that time, my parents were very conservative; they would not open up in front of other family members/us. I learned later that doctors had to remove her breast. But the good thing is, she recovered totally, and it's been 20 years. She has an everyday life. I believe she received good treatment. Every treatment doesn't go wrong!* [b009]

> *Of the few breast-cancer and other cancer survivors I know of, I say it is common for breast cancer patients to feel humiliated or more mentally upset than other cancer patients. …… I don't notice any initiative in our society in this matter; neither is it commonly seen to provide some counseling for the patients for overcoming/fighting cancer.* [b005]

For several reasons, anyone with BC is reluctant to share their experiences. In many cases, patients with BC suffer from depression [17] due to a lack of a support system to debunk the body image and societal myths. The image of an ideal woman propagated in the media (television or social media) reinforces the narrative that is sometimes unhealthy to a person's body image [27]. Particularly the lack of influential media personalities sharing their stories of struggle and recovery or the absence of characters that represent the struggles of patients or survivors of cancer in television shows or media portrayals impacts the perception of the journey and struggles.

Our participants echoed these sentiments in their statements:

> *There is a taboo around breast cancer, regardless of class and status. People do not talk about it. We learn about it later, like once someone has recovered or died of breast cancer. Mostly, no one shares their journey with breast cancer.* [b014]

> *She (a breast cancer survivor) used to talk to me and hang out with our family. However, she and I never*

> *talked about it. Neither I asked, nor she opened up. If it is not very necessary, we will not share. That is the mindset.* [b012]

The term cancer comes with a connotation of impending doom and death; however, the significantly higher chances of patients surviving cancer because of early detection strategies and awareness are not emphasized at all [28]. When it comes to the 5-year overall survival, a study reported it to be 95% for stage I patients, 92% for stage II, 70% for stage III, and only 21% for stage IV patients [29]. It is important to highlight these statistics widely to encourage early detection and awareness. Survival rates of patients with BC are lower compared with Western countries, primarily attributed to early age onset, late-stage detection and delayed initiation of definitive management, and inadequate or fragmented treatment [28].

Second, cultural sensitivities. In addition to not expressing discomfort, awareness campaigns in India and Bangladesh must be wary of the strict cultural, religious, and moral censorship, which does not allow the posting of revealing images of breasts. People, in general, are brought up in an environment where they do not feel comfortable discussing sexual health or women's health concerns, including BC. The term "breast" is perceived through the lens of hypersexualization and hence comes under deep scrutiny from moral and religious sensitivities. As researchers embark on developing awareness campaigns, it becomes crucial to exercise cultural sensitivity, especially considering the conservative nature of the audience in these regions. Being mindful of cultural nuances is essential to ensure that the true essence of the campaign is not overshadowed by moral policing or societal restrictions. It is important to point out that women's health or sexual health and the term breast are not taboo; we cannot undo decades of cultural teaching and biases in a day. Hence respecting the norms and designing awareness campaigns around them would be a more sensitive approach.

> *She (my paternal aunt, who was a teacher) had lung infections and troubled breathing due to breast cancer, but the doctor tracked it back to breast cancer after doing several medical tests; she was never open about the symptoms (or discomforts with her breasts) upfront to the doctor.* [b001]

> *People are not comfortable talking about it. Since I am the son, my mom also never shared about her symptoms or sufferings.* [b007]

Third, women's health care is a necessity, not a luxury. Irrespective of South Asian patriarchal societies or more progressive Western societies, a woman's role in society, economy, and family cannot be ignored. The roles of caregiver, parent, and wife for a rural woman with added burdens of maintaining appearance and body image in the case of an urban woman all contribute to the issues that the contributions women make to the world are "nonproductive" or "nonessential." This perception sometimes from women themselves (in some cases, the family members) leads to the health concerns of women being ignored or underprioritized [23]. This includes a lack of attention to proper nutrition, timely intervention, care after diagnosis, and adequate mental health support. This practice or habit stood out as one of the major reasons why female patients often get a late detection of health-related problems.

> *We wait and hope that the discomforts will disappear automatically after a while – so we tend to linger. I have noticed this among my friends and family (especially females), including myself; we tend not to go to doctors so easily. I am not sure if it's a cultural thing! But it is less about financial conditions and more about awareness.* [b014]

Certainly, that mindset is changing gradually. Nonetheless, there is a need for increased awareness in this particular area:

> *A woman in the family will likely delay her treatment as she thinks it's the least important issue and is probably putting financial pressure on the family. But sometimes, they forget that delayed treatments may invite worse consequences.* [b003]

The Disconnect Between Health Care Models and People

Although none of the existing health care models worldwide are perfect, in the context of a deadly disease like BC and health-related supports, we noted the flaws, inconsistencies, and lack of patient-centered policies in the regions.

First, lack of a yearly health examination policy. Unlike in the United States or other parts of the world, Bengalis usually do not go for yearly health examinations, or it is not a common practice among doctors to prescribe yearly health checkups. The lack of adequate resources, such as available medical professionals, provider-centric consultations, and a symptom-based health care model, is partially responsible for this issue [30]. The region requires immediate advocacy for a robust primary health care system and the use of existing community care efforts to enhance awareness.

Second, out-of-pocket medical expenses. Health examinations are expensive, and it is unusual for the region to support health insurance employers [31]. There is no system of national health insurance in Bangladesh [31]. People whose employers cover health benefits also must pay to the facilities out of pocket first and then later ask for a reimbursement. This cost barrier leads to patients not seeking treatment early on.

Third, lack of integrated service and long-term plans. Those on tight budgets usually have to prioritize the price over reliability and integrated service available to few modern (and more expensive) health care facilities. Patients and their family members are often not fully aware of the overall costs (long-term plans alongside short-term plans) and risk factors so that they can make informed decisions.

> *Talking to the patient and their family about the situation and possibilities is crucial. Only then can we make an informed decision. But if the doctor dictates the options, the patient's family has to follow the order, no matter how expensive or if they have to lose their property in arranging the financial support.* [b009]

Fourth, open conversation between patients and the doctors. Due to various reasons, including high volume of patients,

doctors seem to spend much less time with the patients, which has been a consistent story we heard from our participants.

> *The more renowned the doctors are, the less interactive they are, or the less compassionate they are. There is much less scope and the chance of openly conversing with them.* [b014]

Fifth, insensitive and judgmental commenting. Medical professionals lack adequate training when addressing patients and their concerns. A dearth of empathy compounded by the enormous patient loads leads to negative interactions with the health care providers, and patients or family members do not feel comfortable asking essential questions. An insensitive comment may cause severe damage to the patient's (or their caregiver's) confidence and motivation. Effective doctor-patient communication is determined by the doctor's bedside manner, which patients judge as a major indicator of their doctor's general competence [32,33]. Doctor-patient communication skills are based on verbal and nonverbal communication, where 22% is transferred through voice tone, 55% through visual cues, and 7% verbally [34]. Lack of training to demonstrate effective behavior leads to scarring patient outcomes, one such is listed below:

> *I remember an oncologist's comment to his patient (lung cancer survivor, 65+ years of age) regarding the side effects of Chemo/radiotherapy: you don't have your natural hair anymore! So why worry about that?* [b003]

Sixth, the doctor's gender. BC disproportionately afflicts the female population, and since it concerns an intimate body part, it is common among the patients and their family members to withhold the treatment until they find a female specialist; in short, they prioritize the doctors' gender over professional expertise [32]. In a cross-sectional study conducted with 1257 female patients in Udaipur, India (August 2014), by Nagarajappa et al [32] found an overwhelming 85% (1069/1257) of the respondents said yes to the question: "I would always prefer being treated by a doctor of my gender." In some other similar studies conducted in 2019, researchers found that the most important attributes of maternal health care facility choice for Bangladeshi women were consistent access to a female doctor, the availability of branded drugs, respectful provider attitudes, and a continuum of maternal health care [35,36].

> *My aunt and uncle delayed the treatment as they were hesitant about choosing a doctor; they were looking for a female doctor, but there were not many female breast cancer specialists in Bangladesh during that period. My uncle (my mother's cousin) had hesitation. Had she been diagnosed earlier, she would have had a better chance of survival. She was a survivor, but her entire family suffered greatly due to the delay. My uncle later regretted his decision. I remember him expressing that to my mother (his sister).* [b003]

Although 61.1% (198/324) of medical students in India identify as female, a study by Bajpai et al [37] in 2020 found that men led 67.3% (218/324) of oncology teams in India. These numbers are similar to the numbers in the United States, where out of 667 female respondents, 442 identified as academic oncologists versus nonacademic ones [38]. This gender disbalance among physicians leads to patients not seeking timely medical care.

Seventh, lack of trust. Many people in the region believe that:

> *…an institute (hospital and caregiving center) that is capitalist in nature cannot have people's best interest at heart.* [b016]

This may play a significant role in people's hesitancy to seek health advice unless necessary.

Mental Health Support Infrastructure

All participants expressed concerns about lacking mental health support systems. Most reported that:

- family members and caregivers are never made aware of the patient's mental health, its relevance, and how to handle them [26].
- "I didn't know and never thought mental health could be impacted directly by these (or any) diseases." [b007]
- there is no infrastructure to support the mental health of patients suffering from terminal or deadly diseases.
- "Mental health is considered taboo, more significant of a taboo than breast cancer. People don't usually talk about it. Doctors also don't do a good job in this regard. Providing professional mental health support is not common in India/West Bengal to battle cancer." [b016]
- "Since cancer is associated with a lot of uncertainty, fear, and stress, the moment patients get diagnosed with cancer, they must get counseling. Mental support from proper channels can help patients gather the strength and courage to defeat it." [b010]

Role of Media

Media has always played a significant role in spreading awareness among the general population on various issues. We wanted to learn how our participants felt about the role of media around them in spreading BC awareness. A survey conducted by YouGov Global Media in India identified that in 2022, 59% of Indians under the age of 24 years depended on websites or web apps as their primary media source, whereas people above the age of 45 years depended on television (live or nonlive broadcast) as a primary media source [39]. In Bangladesh, USAID conducted a similar survey and found that 46.6% of respondents mentioned television, and then 33.1% said the internet was a primary source of information media [40]. Similarly, in India, 60% of the population in the age group of 45 years and above depend on television for news, whereas younger audiences tend to lean toward updates from social media instead of watching traditional television news (which broadcasts news at specific hours).

Social media and traditional media (television and newspapers) can substantially spread awareness among people. We wanted to understand whether the media's contribution to raising awareness was limited to a date or time (eg, Breast Cancer Awareness Month) or whether there is any consistent campaigning.

Our participants have reported seeing numerous posts, symbols (such as the pink ribbon or black screen), and talk shows on specific days, weeks, or months of the year (such as during

Breast Cancer Awareness Day). However, these efforts are transitory. They come and go; many posts are left unattended, unread, and unfollowed, and many programs and talk shows are left unwatched. The lack of consistent programming on any platform that regularly sends out relevant messages throughout the year leads to reduced engagement and interest.

> *I have seen many programs aired on the national television of Bangladesh, but I believe that is not sufficient…. I haven't noticed any program regularly on this topic. If we talk about social media, I don't see enough posts relevant to this topic.* [b004]

> *People share (even if that's at a minimal level, especially on special occasions like Breast Cancer Month). But others may not have time to read it carefully due to their busy schedules. I sometimes see such posts but don't have time to read them thoroughly. I feel until it hits someone, they are likely to ignore (or blindside) it.* [b003]

Most BC-related posts (eg, Figure 4) on media platforms include either a pink bow [41,42] or wearing the color pink at campaign events [43,44]. To urban audiences' mammograms are widely promoted by health care professionals or organizations [41,45].

Figure 4. Media coverage, rallies, and campaigns.

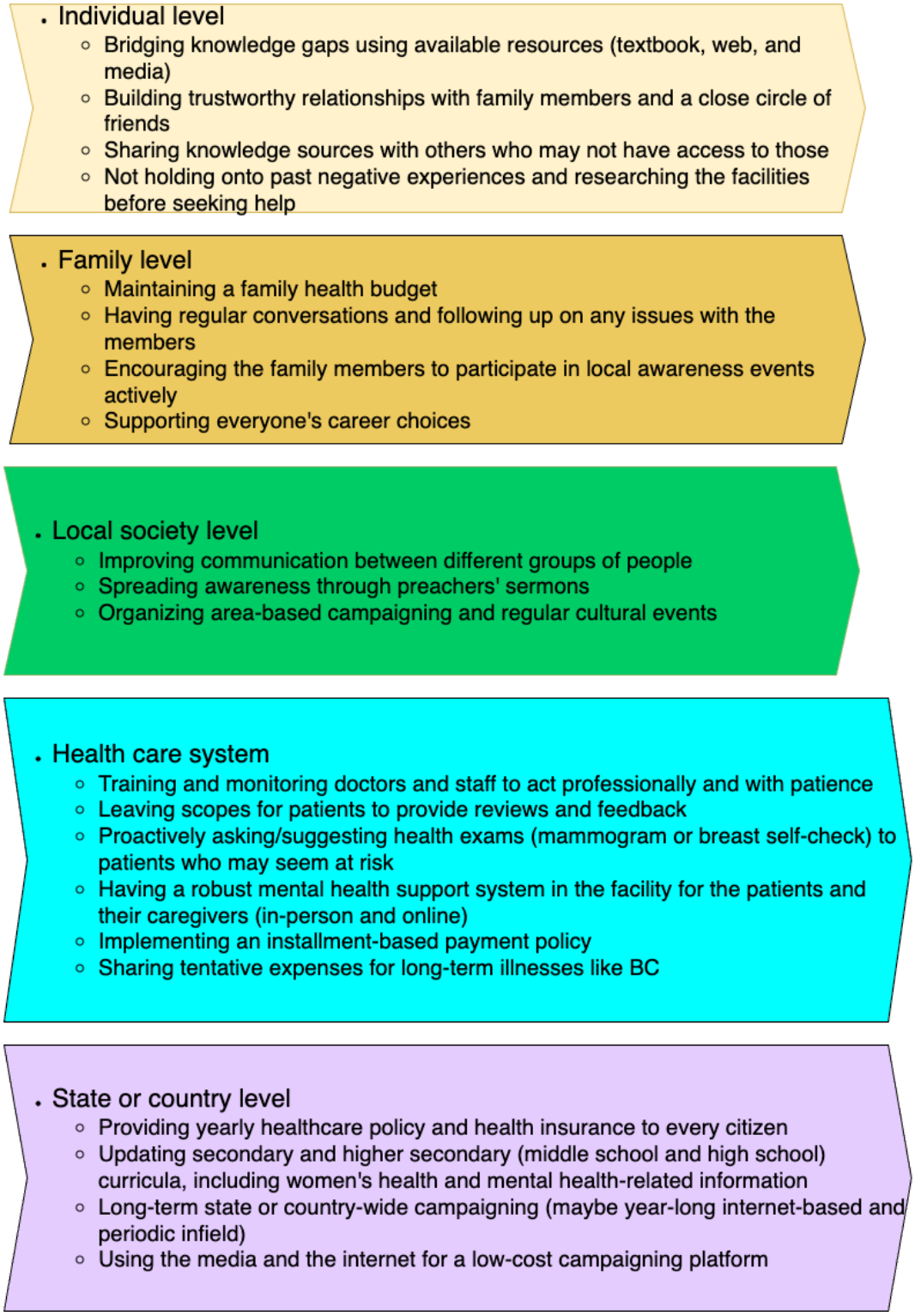
The absence of relevant information about the disease, prognosis, and symptoms is stark and leaves the audience wondering how to perceive this awareness campaign. Often, momentum is lost, and true engagement is not achieved, leading to lackluster awareness numbers.

> *We see advertisements relating to breast cancer, sometimes on the television or elsewhere, or billboards, but then it is very short and crisp. And it just says that we should be aware of breast cancer, but then what to be aware of it is never mentioned anywhere.* [b017]

Awareness Raising

Many of the findings are codependent and mutually recursive. For example, lack of trust in health care systems in the area and financial hardships implicitly push people to neglect early symptoms of deadly diseases; lack of open conversation and social stigma block them from reaching out for mental health support, to name a few. Our discussions with the participants highlighted several means to help spread mass awareness in the regions. In this section, we present the summary of our suggestions to achieve mass awareness through Figure 5 and briefly discuss them right after.

Figure 5. Our recommendation to help raise awareness. BC: breast cancer.



Designing a Yearly Health Care Routine for Everyone

> *Society must believe that early diagnosis can help increase survival and reduce suffering. We need to build the practice of going through a yearly checkup.* [b008]

Given the population, unemployment, and other issues, achieving mass success in building this practice will require support and cooperation from international organizations (such as the World Health Organization) and local governments.

Spreading the Voice of Survivors

We often hear words from specialists (doctors and scientists) on mass media (eg, Shastho Kotha [46]). Aside from their knowledgeable notes, the programs will draw attention to the mass scale if BC survivors' interviews are included regularly. This will also reduce fear, hesitancy, and social stigma. When

people see interviews of BC survivors belonging to the same community (age, gender, social status, and so on), it will help everyone to open up about similar issues.

> *I believe cancer changes our views and ways of life. We must break the practice of neglecting minor symptoms and not seeking medical advice. We need to improve trust and healthy communication between doctors and patients. Roughly 50% of the population is female. So, whichever cancer is more probabilistic to which population, they need to take it seriously. Financial hurdles should not be the first concern. It will be beneficial if there is a platform where people can talk about it and not treat it as something that cannot be discussed.* [b016]

Sharing the experiences of BC survivors as they navigate treatment plans, manage financial burdens, and embark on their journey toward recovery can play a crucial role in dispelling fear, reducing anxiety, and challenging the taboos surrounding BC. Narratives that illuminate the resilience of these survivors, highlight the support of their families, and shed light on the obstacles they overcome contribute to humanizing this daunting illness.

Using the Grassroots Health Workers

> *NGOs like Surjer Hashi Clinic, Shobuj Chata Community Clinic, etc., wildly succeeded in child immunization and birth control in Bangladesh. So, if cancer awareness (like breast and cervical cancer) is taken as an agenda by such organizations, I am hopeful it will succeed.* [b003]

If the health workers spread a simple flyer about BC with some visuals and depictions in Bangla during their weekly visits, it will help raise awareness in rural areas quickly.

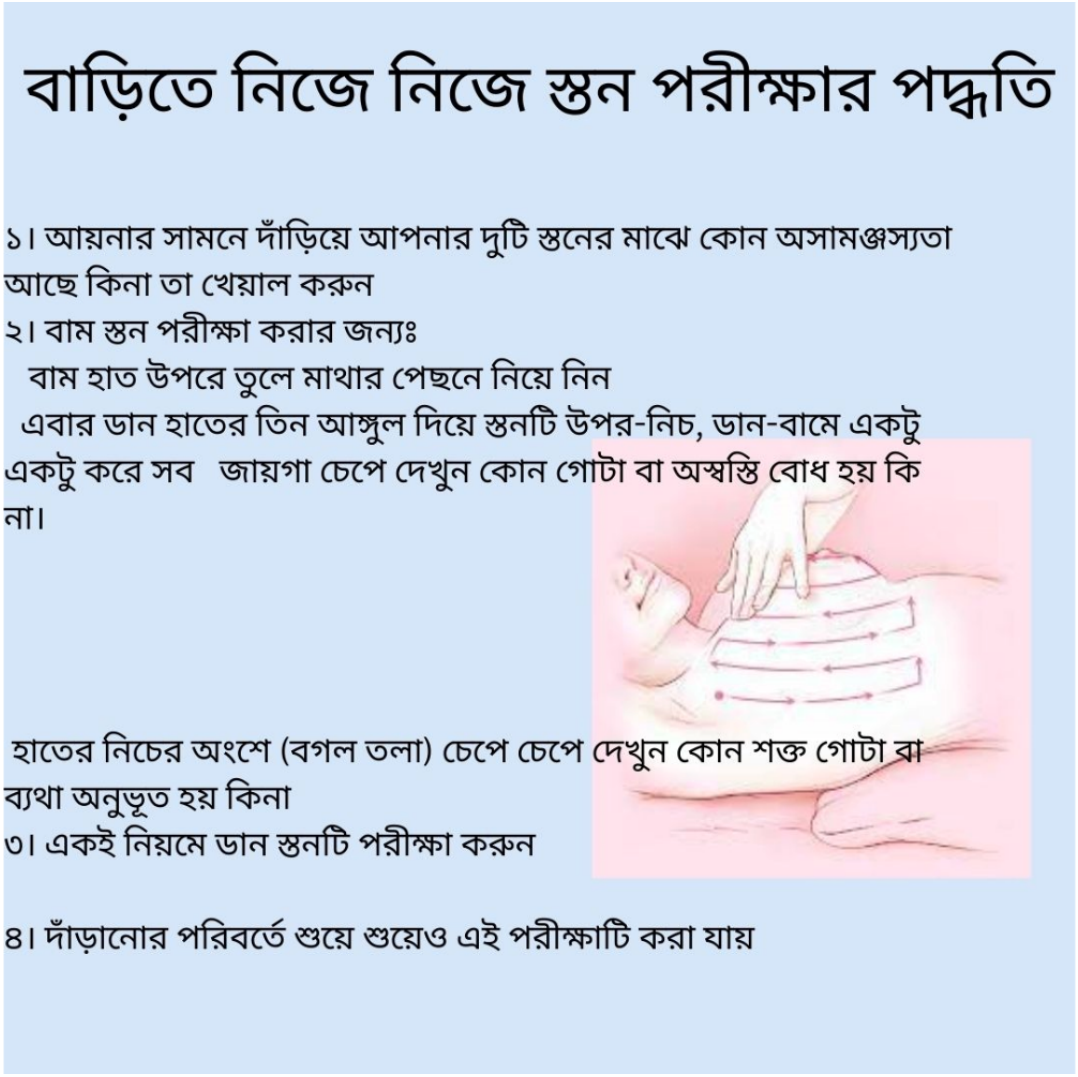
> *Bangladesh has successfully rolled out the idea of condoms and contraception, though we are considered a very conservative society. Major religions of our land directly oppose the idea of birth control. Yet the success came not because of the paid advertisements or programs in media but because of sending medical advice to the grassroots level. So, I believe we can succeed in growing awareness about breast cancer if the Government takes such programs into the hand.* [b005]

Establishing a Collaborative Effort on Social Media

Bangladesh's internet penetration rate stood at 31.5% (52.58 million/167.1 million) of the total population at the start of 2022 [47], and India's 47% (658 million/1.40 billion) of the population has internet access [48]. India and Bangladesh have both shown steady growth in mobile cell phone service, too. Instant messaging apps such as WhatsApp (Meta; 487 million users in India and 40 million users in Bangladesh) [49,50]. We strongly believe these platforms can be used to spread accurate awareness messages to users, particularly if government health organizations partner.

We suggest regularly preparing WhatsApp message broadcasts or mass forwards to groups, patients, or individuals from credible organizations such as local government branches (eg, zilla panchayat or district offices), medical associations (eg, Bangladesh Medical Association or Medical Council of India), or nonprofit organizations (eg, Sabuj Sathi). Sending these messages individually to users also avoids making anyone uncomfortable with the content, keeping in mind the social and religious sentiments or barriers in these regions. WhatsApp's ubiquitous presence in the digital communication ecosystem of both countries removes additional app downloads or installations. We propose using WhatsApp forwards (illustrated in Figure 6) as a cost-effective, easy-to-understand, and time-efficient method for dissemination. However, it is essential that these messages are reviewed by an expert before distribution to ensure accuracy and reliability.

Figure 6. Proposed self-breast examination instruction flyer.



Inclusive Texts in Secondary and Higher Secondary (Middle and High School) Curricula

Many people in both regions finish their education till middle- or high-school level. Some inclusive articles (written with appropriate languages and cultural values in mind) regarding women's health and the terminal and most predominant diseases in society will help spread awareness. The students and their families will gradually become aware; thus, many of the causes blocking the attention can be addressed thus far. Because textbooks can reach families where social media or entertainment media cannot.

> *I grew up in a village. I can talk to you from that context.… It isn't easy to spread awareness in the villages. For example, I grew up in a Muslim family where watching TV was not allowed, as it is anticipated that kids will be misguided by watching cartoons or inappropriate programs. So, getting this knowledge through TV was/is challenging for many kids like me.* [b006]

Per our research, no story, article, or novel about BC is included in the secondary and higher-secondary curricula of Bangladesh and West Bengal, India.

Educating the Caregivers and the Family Members

Besides building infrastructure for mental health support, doctors and other associates must educate the family members and caregivers about the patient's mental health and how to deal with it. We heard stories suggesting how a family can motivate a survivor to fight BC:

> *Someone very close to me has breast cancer. Initially, when it was diagnosed, the doctor prescribed her some medicine but barely communicated with her about her mental health. So, she melted down. After battling breast cancer for two years, she started self-motivating and regained her mental strength. She is a mother of two young kids, so that pushed her to keep going and not give up. She involved herself in activities like Yoga. Due to regaining mental strength, she started handling the treatment better (I am not suggesting that the cancer has disappeared, but she started driving it better).* [b002]

> *She (a relative) went into depression after the mastectomy. But since she has kids, she motivated herself. Her family also supported her very strongly. This support has helped her come out of the trauma. She is still following up with her breast cancer-related*

treatment. It is four years that she has been receiving treatments. [b011]

Discussion

In this research study, we aimed to understand the attitudes and perceptions of South Asian people, specifically people from Bangladesh and West Bengal, India, regarding BC. Our goal was to identify common barriers to raising mass awareness about BC in this population. In the results section, we discussed the significant barriers to BC awareness, which included a lack of knowledge and open conversation, negligence of early symptoms, negative bias toward health care systems, judgmental attitudes, and insensitive comments toward patients and their families, and financial hardship with limited options for financial support. We also provided suggestions to effectively raise mass awareness, including encouraging open conversations among family members and different groups, discussing relevant topics in social events and places of religious practices, integrating the topic into school curriculums, and disseminating accurate information through social and traditional media.

We present, in this work, the summaries of our findings in Figure 2 and Figure 5, where we clustered the critical points into 5 themes based on different societal levels—individual, family, local society, health care system, and state or country. Our findings highlighted the need for long-term planning and cooperation between people and the government, collaboration between national and international organizations, the use of media and education systems, and the importance of modern technologies and mass media in raising BC awareness.

Considering our dataset size (N=17) and the demographics of the participants, particularly their socioeconomic status and educational background, it is important to acknowledge that our findings are not fully representative of the entire population. Specifically, our dataset may not adequately capture the experiences of individuals living in rural areas, those below the poverty line, or those in the wealthiest brackets. These socioeconomic statuses certainly have disparate health care experiences and struggles. Our participant pool does offer valuable insights into the perspectives of the upper-middle-class segment of society and gives a general overview of barriers that should be further investigated.

For future work, it will be essential to extend this study to include participants from a wider range of economic and educational backgrounds. This broader approach will enable us to evaluate the generalizability of our findings and recommendations across diverse demographic groups and to uncover any critical issues that may have been overlooked in the current analysis. Such an expansion will contribute to a more comprehensive understanding of the nuances of this important issue and increase the applicability of our conclusions to a broader audience.

Conflicts of Interest

None declared.

References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-249 [FREE Full text] [doi: 10.3322/caac.21660] [Medline: 33538338]
2. Breast Cancer.: World Health Organization; 2024. URL: https://www.who.int/news-room/fact-sheets/detail/breast-cancer [accessed 2024-11-27]
3. Home | Government of India --- censusindia.gov.in? Government of India. URL: https://censusindia.gov.in/census.website/ [accessed 2023-07-26]
4. Press Information Bureau. Average Life Expectancy --- pib.gov.in,? Government of India. URL: https://pib.gov.in/PressReleseDetail.aspx?PRID=1606209 [accessed 2023-10-08]
5. Bangladesh Life Expectancy 1950-2023. URL: https://www.macrotrends.net/countries/BGD/bangladesh/life-expectancy [accessed 2023-07-31]
6. West Bengal Population| Sex Ratio| Literacy rate. 2023. URL: https://www.census2011.co.in/census/state/west+bengal.html [accessed 2023-10-19]
7. Leong SPL, Shen ZZ, Liu TJ, Agarwal G, Tajima T, Paik NS, et al. Is breast cancer the same disease in Asian and Western countries? World J Surg 2010;34(10):2308-2324 [FREE Full text] [doi: 10.1007/s00268-010-0683-1] [Medline: 20607258]
8. Hossain MS, Ferdous S, Karim-Kos HE. Breast cancer in South Asia: a Bangladeshi perspective. Cancer Epidemiol 2014;38(5):465-470. [doi: 10.1016/j.canep.2014.08.004] [Medline: 25182670]
9. Islam MR. Urbanization in Bangladesh: The prevalence of breast cancer brings unique challenges. 2021. URL: https://ascopost.com/issues/october-25-2021/urbanization-in-bangladesh-the-prevalence-of-breast-cancer-brings-unique-challenges/ [accessed 2023-10-08]
10. INDIA: BREAST CANCER. www.worldlifeexpectancy.com. 2024. URL: https://www.worldlifeexpectancy.com/india-breast-cancer [accessed 2024-11-27]
11. BANGLADESH: BREAST CANCER. worldlifeexpectancy.com. 2020. URL: https://www.worldlifeexpectancy.com/bangladesh-breast-cancer [accessed 2024-11-27]

12. Naderifar M, Goli H, Ghaljaie F. Snowball sampling: a purposeful method of sampling in qualitative research. Strides Dev Med Educ 2017;14(3). [doi: 10.5812/sdme.67670]
13. Jenks AC. From "lists of traits" to "open-mindedness": emerging issues in cultural competence education. Cult Med Psychiatry 2011;35(2):209-235. [doi: 10.1007/s11013-011-9212-4] [Medline: 21560030]
14. Bingham AJ, Witkowsky P. Deductive and inductive approaches to qualitative data analysis. In: Analyzing and Interpreting Qualitative Data: After the Interview. Washington DC: Sage; 2021:133-146.
15. Mattimoe R, Hayden MT, Murphy B, Ballantine J. Approaches to analysis of qualitative research data: a reflection on the manual and technological approaches. Accounting, Finance & Governance Review 2021;27(1):54-69 [FREE Full text]
16. Haque M, Kawsar M, Adibuzzaman S, Ahamed S, Love R, Dowla R, et al. Findings of e-ESAS: a mobile based symptom monitoring system for breast cancer patients in rural Bangladesh. 2012 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2012 May 5-10; Austin Texas USA. [doi: 10.1145/2207676.2208532]
17. Amin MN, Uddin MG, Uddin MN, Rahaman MZ, Siddiqui SA, Hossain MS, et al. A hospital based survey to evaluate knowledge, awareness and perceived barriers regarding breast cancer screening among females in Bangladesh. Heliyon 2020;6(4):e03753 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03753] [Medline: 32382673]
18. Statistics of Breast Cancer in India. URL: https://cytecare.com/blog/breast-cancer/statistics-of-breast-cancer/ [accessed 2023-07-26]
19. Malvia S, Bagadi SA, Dubey US, Saxena S. Epidemiology of breast cancer in Indian women. Asia Pac J Clin Oncol 2017;13(4):289-295. [doi: 10.1111/ajco.12661] [Medline: 28181405]
20. Breast Cancer Risk in American Women. URL: https://www.cancer.gov/types/breast/risk-fact-sheet [accessed 2023-07-28]
21. Breast Cancer Epidemiology and Risk Factors: Epidemiology, Reproductive Factors, Endogenous Sex Hormones and Risk of Breast Cancer. URL: https://emedicine.medscape.com/article/1697353-overview [accessed 2023-07-26]
22. Dodge DG, Jarowenko AM. Diagnosed with breast cancer: stigmatized or member of an empowered sisterhood? In: Stigma and Prejudice: Touchstones in Understanding Diversity in Healthcare. Cham: Springer International Publishing; 2016:145-166.
23. Mansoor T, Abid S. Negotiating femininity, motherhood and beauty: experiences of Pakistani women breast cancer patients. Asian Journal of Women's Studies 2020;26(4):485-502. [doi: 10.1080/12259276.2020.1859076]
24. Agha N, Tarar MG. Battling breast cancer: women’s narratives of struggle, family support and survival from rural Sindh, Pakistan. Pak. J. Women Stud 2019;26(1):21-40 [FREE Full text]
25. McClelland SI. Gender and sexual labor near the end of life: advanced breast cancer and femininity norms. Women's Reprod Health 2017;4(1):29-45. [doi: 10.1080/23293691.2017.1276367]
26. Daniel S, Venkateswaran C, Hutchinson A, Johnson MJ. 'I don't talk about my distress to others; I feel that I have to suffer my problems...' voices of indian women with breast cancer: a qualitative interview study. Support Care Cancer 2021;29(5):2591-2600 [FREE Full text] [doi: 10.1007/s00520-020-05756-8] [Medline: 32955655]
27. Nagar I, Virk R. The struggle between the real and ideal: impact of acute media exposure on body image of young Indian women. Sage Open 2017;7(1):215824401769132. [doi: 10.1177/2158244017691327]
28. Mehrotra R, Yadav K. Breast cancer in India: present scenario and the challenges ahead. World J Clin Oncol 2022;13(3):209-218 [FREE Full text] [doi: 10.5306/wjco.v13.i3.209] [Medline: 35433294]
29. Maurya AP, Brahmachari S. Current status of breast cancer management in India. Indian J Surg 2020;83(6):316-321. [doi: 10.1007/s12262-020-02388-4]
30. van Weel C, Kassai R, Qidwai W, Kumar R, Bala K, Prasad Gupta P, et al. Primary healthcare policy implementation in South Asia. BMJ Glob Health 2016;1(2):e000057 [FREE Full text] [doi: 10.1136/bmjgh-2016-000057] [Medline: 28588938]
31. Story HL, Love RR, Salim R, Roberto AJ, Krieger JL, Ginsburg OM. Improving outcomes from breast cancer in a low-income country: lessons from Bangladesh. Int J Breast Cancer 2012;2012:423562 [FREE Full text] [doi: 10.1155/2012/423562] [Medline: 22295245]
32. Nagarajappa R, Bapat S, Asawa K, Daryani H, Gupta VV, Chaturvedi P, et al. Perception of sexual ramification among women toward male doctors in today's Indian society. Transl Behav Med 2018;8(5):771-775. [doi: 10.1093/tbm/ibx027] [Medline: 29385565]
33. Hall JA, Roter DL, Rand CS. Communication of affect between patient and physician. J Health Soc Behav 1981;22(1):18-30. [doi: 10.2307/2136365]
34. van den Brink-Muinen A, Verhaak PFM, Bensing JM, Bahrs O, Deveugele M, Gask L, et al. Communication in general practice: differences between European countries. Fam Pract 2003;20(4):478-485. [doi: 10.1093/fampra/cmg426] [Medline: 12876125]
35. Mahumud RA, Alamgir NI, Hossain MT, Baruwa E, Sultana M, Gow J, et al. Women's preferences for maternal healthcare services in Bangladesh: evidence from a discrete choice experiment. J Clin Med 2019;8(2):132 [FREE Full text] [doi: 10.3390/jcm8020132] [Medline: 30678044]
36. Begum SA, Mahmud T, Rahman T, Zannat J, Khatun F, Nahar K, et al. Knowledge, attitude and practice of Bangladeshi women towards breast cancer: a cross sectional study. Mymensingh Med J 2019;28(1):96-104. [Medline: 30755557]
37. Bajpai J, Mailankody S, Nair R, Surappa ST, Gupta S, Prabhash K, et al. Gender climate in Indian oncology: national survey report. ESMO Open 2020;5(2):e000671 [FREE Full text] [doi: 10.1136/esmoopen-2020-000671] [Medline: 32354723]

38. Oncologists in Mumbai Address Elephant in Room: Gender Gap Among Cancer Specialists. URL: https://timesofindia.indiatimes.com/city/mumbai/oncologists-address-elephant-in-room-gender-gap-among-cancer-specialists/articleshow/86761995.cms [accessed 2023-07-26]
39. Bureau M. Digital media dominates consumers' choice in India but growth in consumption appears to be slowing: Report. 2022. URL: https://www.medianews4u.com/digital-media-dominates-consumers-choice-in-india-but-growth-in-consumption-appears-to-be-slowing-report/ [accessed 2023-07-27]
40. Final Report, Assessment of the Media Sector in Bangladesh. USAID. 2022. URL: https://pdf.usaid.gov/pdf_docs/PA00Z9T6.pdf [accessed 2024-11-27]
41. Check-up camp for breast cancer in Lucknow. Lucknow News. 2023. URL: https://timesofindia.indiatimes.com/city/lucknow/check-up-camp-for-breast-cancer-in-lucknow/articleshow/98055384.cms [accessed 2023-07-27]
42. Breast Cancer Awareness Campaign for Early Detection of Cancer. 2022. URL: https://financialsamachar.com/breast-cancer-awareness-campaign-for-early-detection-of-cancer/ [accessed 2023-07-27]
43. Fresh Tissue observes 'Pink Day' on the occasion of breast cancer awareness month. 2022. URL: https://www.tbsnews.net/economy/corporates/fresh-tissue-observes-pink-day-occasion-breast-cancer-awareness-month-516138 [accessed 2023-07-28]
44. Shifting Tide: Breast Cancer is Number 1 Killer of Women. 2026. URL: https://timesofindia.indiatimes.com/home/science/shifting-tide-breast-cancer-is-number-1-killer-of-women/articleshow/51107576.cms [accessed 2023-07-27]
45. Ruby General Hospital YouTube page. Breast Cancer Awareness From Chandan Sen, Television Actor. 2021. URL: https://www.youtube.com/watch?v=OKfG0i2ZLM0 [accessed 2023-07-27]
46. BanglaVision PROGRAM YouTube page. Shastho Kotha| Ep 27| Tahamina Mukta| BanglaVision Program 2019. 2019. URL: https://www.youtube.com/watch?v=twxY5NVrto0 [accessed 2023-10-19]
47. Kemp S. Digital 2022: Bangladesh. Data Reportal Global Digital Insights. URL: https://datareportal.com/reports/digital-2022-bangladesh [accessed 2023-07-31]
48. Kemp S. Digital 2022: India. Data Reportal Global Digital Insights. URL: https://datareportal.com/reports/digital-2022-india [accessed 2023-07-31]
49. Number of WhatsApp users in selected countries worldwide as of June 2021. Statista. URL: https://www.statista.com/statistics/289778/countries-with-the-most-facebook-users/ [accessed 2023-07-31]
50. Solutions O. Ridesharing App OBHAI On WhatsApp - A First in Bangladesh. URL: https://www.obhai.com/blog/ride-with-obhai-through-whatsapp.html [accessed 2023-07-31]

Abbreviations

BC: breast cancer
CT: computed tomography
GLOBOCAN: Global Cancer Observatory
USAID: United States Agency for International Development

Edited by A Kushniruk, Y Gong; submitted 25.10.23; peer-reviewed by S Chaudhry, J Berian, DW Han; comments to author 04.03.24; revised version received 31.05.24; accepted 08.11.24; published 10.01.25.

Please cite as:
Hamid F, Roy T
Unveiling Sociocultural Barriers to Breast Cancer Awareness Among the South Asian Population: Case Study of Bangladesh and West Bengal, India
JMIR Hum Factors 2025;12:e53969
URL: https://humanfactors.jmir.org/2025/1/e53969
doi:10.2196/53969
PMID:

Original Paper

The Significance of a Cerebrovascular Accident Outcome Prediction Model for Patients, Family Members, and Health Care Professionals: Qualitative Evaluation Study

Corinne G Allaart[1,2], MSc; Sanne van Houwelingen[1,2], MSc; Pieter HE Hilkens[2], MD; Aart van Halteren[1,3], PhD; Douwe H Biesma[4], MD, PhD; Lea Dijksman[2], PhD; Paul B van der Nat[2,5,6], PhD

[1]Department of Computer Science, Faculty of Science, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[2]Department of Value Improvement, St. Antonius Hospital, Nieuwegein, Netherlands
[3]Philips Research, Eindhoven, Netherlands
[4]Leiden Unversity Medical Center, Leiden, Netherlands
[5]IQ Healthcare, Radboud University Medical Center, Nijmegen, Netherlands
[6]Santeon, Utrecht, Netherlands

Corresponding Author:
Corinne G Allaart, MSc
Department of Computer Science
Faculty of Science
Vrije Universiteit Amsterdam
Boelelaan 1111
Amsterdam, 1081 HV
Netherlands
Phone: 31 20 5989898
Email: c.g.allaart@vu.nl

Abstract

Background: Patients with cerebrovascular accident (CVA) should be involved in setting their rehabilitation goals. A personalized prediction of CVA outcomes would allow care professionals to better inform patients and informal caregivers. Several accurate prediction models have been created, but acceptance and proper implementation of the models are prerequisites for model adoption.

Objective: This study aimed to assess the added value of a prediction model for long-term outcomes of rehabilitation after CVA and evaluate how it can best be displayed, implemented, and integrated into the care process.

Methods: We designed a mock-up version, including visualizations, based on our recently developed prediction model. We conducted focus groups with CVA patients and informal caregivers, and separate focus groups with health care professionals (HCPs). Their opinions on the current information management and the model were analyzed using a thematic analysis approach. Lastly, a Measurement Instrument for Determinants of Innovations (MIDI) questionnaire was used to collect insights into the prediction model and visualizations with HCPs.

Results: The analysis of 6 focus groups, with 9 patients, 4 informal caregivers, and 8 HCPs, resulted in 10 themes in 3 categories: evaluation of the current care process (information absorption, expectations of rehabilitation, prediction of outcomes, and decision aid), content of the prediction model (reliability, relevance, and influence on the care process), and accessibility of the model (ease of understanding, model type preference, and moment of use). We extracted recommendations for the prediction model and visualizations. The results of the questionnaire survey (9 responses, 56% response rate) underscored the themes of the focus groups.

Conclusions: There is a need for the use of a prediction model to assess CVA outcomes, as indicated by the general approval of participants in both the focus groups and the questionnaire survey. We recommend that the prediction model be geared toward HCPs, as they can provide the context necessary for patients and informal caregivers. Good reliability and relevance of the prediction model will be essential for its wide adoption.

(JMIR Hum Factors 2025;12:e56521) doi:10.2196/56521

KEYWORDS

cerebrovascular accident; machine learning; artificial intelligence; visualization; focus groups; questionnaire; informal caregivers; health care professionals

Introduction

Cerebrovascular accident (CVA), the collective name for ischemic stroke and intracerebral hemorrhagic stroke, affects 15 million people each year globally [1]. A CVA can have a range of symptoms that vary from patient to patient. Unilateral paralysis or weakness, numbness or loss of vision, speech difficulties and other cognitive dysfunctions, ataxia, diplopia, and nonorthostatic dizziness are some of these symptoms [2]. After the acute phase of treatment, patients often undergo a long rehabilitation process, although the length, intensity, and outcome can vary greatly between patients. In the Netherlands, 65% of CVA patients undergo rehabilitation at home after their hospitalization. The remaining 35% of CVA patients visit a rehabilitation center or geriatric rehabilitation center. Among patients in this group, 25% will not return to their homes and will go to nursing homes [3].

To provide the best care, health care professionals (HCPs) and patients should decide together which path to take in the care process [4]: shared decision-making. Efficient and practical goals can be determined when patients are involved in decision-making, leading to a more appropriate and suitable discharge location. Thus, CVA patients should be involved in decision-making regarding where rehabilitation will take place and setting rehabilitation goals [5]. A considerable number of CVA patients prefer active involvement in decision-making and consider themselves capable of doing so [6]. An important part of supporting active involvement in the decision-making process is to offer patients insights into expected outcomes.

Recovery is very heterogeneous for CVA patients in terms of speed, process, and outcomes, and is often considered too complex for HCPs to predict. As such, little information about expected medium- and long-term outcomes (3 months to >1 year) is provided to patients. Over the last few years, using artificial intelligence (AI) methods, several outcome prediction models have been developed. Some models on certain aspects of functional status have been developed, such as rehabilitation of motoric upper limb skills [7] or independent walking [8]. Others have reported testing at rehabilitation admission [9] or the use of wearable sensors [10] to provide useful data to predict postrehabilitation CVA outcomes. Most of these models use machine learning algorithms on structured clinical data, with varying accuracy [11,12]. Deep learning has offered the opportunity to combine more data and more different data types, such as imaging or free text, into a prediction model [13] and as such increase the accuracy of predictions. Several studies have investigated how deep learning can lead to better CVA outcome predictions [13,14] to reach clinically acceptable reliability. One of our previous studies [15] showed that combining structured clinical data and perfusion computed tomography (CT) scans can lead to an increase in performance.

While having accurate and reliable predictions is important, the explainability and clinical relevance of AI methods are essential for their adoption in medical practice [16,17], and many prediction model implementations currently fall short in these aspects [18]. Using a functional tool based on AI in a real-world setting would require both HCPs and patients to understand and trust the outcome prediction and see its relevance. Understanding a predicted outcome can be challenging for healthy people, but even more for CVA patients. These patients can have impaired cognition and often have a higher age [2]. Thus, information should be provided in an easy and clear manner for low cognitive effort. For example, when numerical data are presented, it is recommended to use simple visual forms such as flow or bar charts and circle diagrams [19].

The discharge interview usually addresses the duration and location of the rehabilitation. There are typically multiple possibilities after discharge, such as going home without or with home therapy, outpatient rehabilitation, geriatric rehabilitation, and clinical rehabilitation treatment [3]. A personalized prediction of CVA outcomes and the length of the rehabilitation process would allow HCPs to better inform patients and informal caregivers. This could form a basis for shared decision-making, as a more informed conversation about the preferences of the patient can be held. This research aims to assess the added value of a prediction model for long-term outcomes of rehabilitation after CVA, to investigate how its results can be best displayed, and to determine the best way to implement and integrate it into the care process.

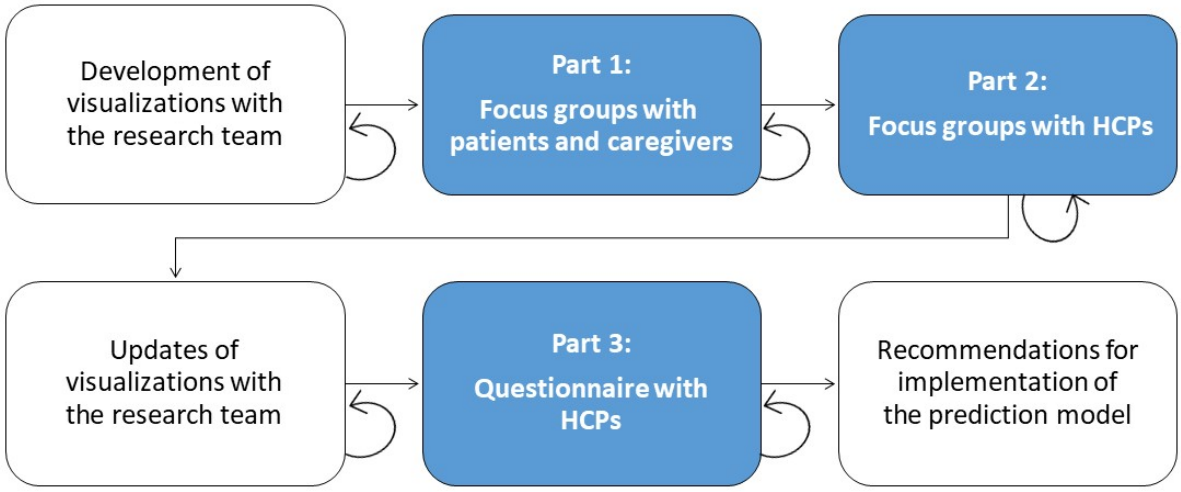
Methods

Design

A qualitative study was designed with the goal of evaluating the added value of a prediction model for stroke outcomes in CVA care. More specifically, this study aims to investigate how to best display the outcomes of the model, and integrate and implement them in the care process in an appropriate manner. We based our study on a prediction model previously designed in our hospital [15]. Before the qualitative evaluation, the research team developed a mock-up of a display of the prediction model, together with several different visualizations of the model outcome and interpretability. The research team included a neurologist and an expert on medical AI. The mock-up design was iterative, allowing for an update after each part of the study.

The study consisted of 3 parts: focus group discussions with CVA patients and their informal caregivers, focus group discussions with HCPs, and a follow-up questionnaire to HCPs, as shown in Figure 1. In both sets of focus groups, the opinions of patients and HCPs on the current information management and their needs with regard to the prediction model were explored and analyzed using a thematic analysis approach. The questionnaire was used to evaluate the finalized prediction model and visualizations with the HCPs.

Figure 1. Schematic overview of the different parts of the study. The circular arrow refers to the iterative aspect of each step. HCP: health care professional.



Focus Groups

Study Population

Patients who were admitted to St. Antonius Hospital between January 2020 and December 2021 for a CVA were selected using convenience sampling. Patients who had a CVA in the previous 3 months or who did not speak Dutch were excluded. We selected participants with a wide variation in age, gender, rehabilitation type (medical specialist, geriatric, home, or outpatient), and education level. An overview of the characteristics of the selected participants can be found in Table 1. Patients were approached by phone, and information letters were sent by email or postal address. Patients could bring their informal caregivers to the focus groups if this was of added value due to impaired cognition or higher age. The research team thought that an informal caregiver could assist in interpreting the model and could discuss patient preferences. Another advantage of including informal caregivers is that they are often involved in the entire hospitalization process of patients, and their opinions can therefore be valuable. Participants signed up by contacting the researcher and were asked to sign an informed consent form before the focus group started.

Table 1. Cerebrovascular accident participants in focus groups.

Participant	Age range (years)	Gender	Level of education	Rehabilitation
P1	71-75	Female	Vocational education	Geriatric rehabilitation center
P2	41-45	Male	Vocational education	Rehabilitation at home
P3	41-45	Female	Vocational education	Rehabilitation center
P4	71-75	Male	Applied university	Rehabilitation at home
P5	81-85	Female	High school	Rehabilitation at home
P6	66-70	Female	Applied university	Rehabilitation at home
P7	41-45	Male	High school	Rehabilitation at home
P8	76-80	Female	Vocational education	Geriatric rehabilitation center
P9	61-65	Female	Vocational education	Rehabilitation center

For HCPs, participants were recruited from a pool of HCPs involved in the discharge or rehabilitation process in the hospital. This included neurologists, rehabilitation doctors, junior doctors, nurses, and nurse specialists. All participants were approached via email or in person. The research team aimed to have 3 participants in each focus group, specifically a medical specialist, a general doctor, and a nurse.

Data Collection

The focus groups were led by 2 researchers (CGA and SvH). SvH moderated the focus groups with patients, and CGA moderated the focus groups with HCPs. We opted for the separation of patients and HCPs in different focus groups, such that the participants in both groups would be able to speak more freely about their experiences. We started with 3 focus groups for patients and their informal caregivers and 3 focus groups for HCPs, and we assessed whether saturation of data was reached or more focus groups were necessary. Each focus group had different participants.

The focus groups consisted of 2 phases. The first phase of the focus group was intended to explore the needs and expectations of the use of a prediction model during the discharge interview. This started with a reflection on the discharge interview and rehabilitation process, what expectations were created, what shared decisions were made, how the process went, and other

questions about topics and components that were addressed in the prediction model. The goal was also to determine how a prediction model could contribute to better expectations and improve the discharge interview.

In the second phase of the focus groups, different options for visualizations of the prediction model were discussed. We have elaborated on their design in the section on visualizations. Before the second phase of the focus group started, there was a short explanation of the visualizations. After this, the participants were asked to discuss several topics. First, whether the visualizations were clearly explained. Second, they were asked about relevance and what they think about the visualizations being used. During the focus groups with HCPs, this also included the predictors included in the model. Finally, they were asked to discuss which of the visualizations they preferred and why.

Focus Group Guide

Focus group guides were used, consisting of open and follow-up questions for the 2 phases of the focus groups (Multimedia Appendix 1 for patients and Multimedia Appendix 2 for HCPs). The focus group guides consisted of a few open questions, aimed at starting a discussion on information provision to patients in the current care process and the suitability of the prediction model and its visualizations. The focus groups followed an iterative design, which meant that after each focus group, the questions in the interview guide were adjusted if certain topics needed to be added or changed.

Data Analysis

Analysis of the focus groups was performed by researchers CGA and SvH in a thematic analysis approach using Atlas.ti [20]. CGA and SvH are early stage researchers with educational training and some research experience in focus group research. The thematic analysis was performed as follows. First, the researchers transcribed and anonymized the interviews. Then, both researchers separately open coded the transcripts. The data were coded axially by clustering codes by meaning, and the clusters were assigned themes. When all 6 transcripts were axially coded, the data were selectively coded to determine the relevance and consistency of the themes [21]. Finally, the themes were defined by noting the characteristics of the themes for consistent use. After analysis, the researchers held discussions about the data until a consensus was reached. We aimed for research triangulation by having 2 researchers independently analyze the data and iterate until agreement.

Questionnaire Survey

Study Population

The study population of the questionnaire consisted of HCPs. Similar to the focus groups, the participants were recruited from the pool of HCPs involved in the discharge or rehabilitation process in the hospital. However, only HCPs who directly worked with the prediction mode, such as junior doctors and neurologists, were asked to fill out the questionnaire. They were contacted by email or phone, or in person.

Data Collection

The last part of this study consisted of a questionnaire for a final analysis of the visualizations by HCPs. A simulated visualization (updated after the focus groups; see section visualizations) was shown to the HCPs for different scenarios for possible situations of patients in a discharge interview. They were asked to answer a modified Measurement Instrument for Determinants of Innovations (MIDI) questionnaire [22], which is a tool for describing the helpfulness and possibilities of medical technological devices. We selected 2 relevant sections out of 4 in the questionnaire regarding the innovation itself and the end users. The other 2 sections regarding work environment and sociopolitical environment were out of scope and more suitable for a later implementation phase. The questionnaire was adapted to fit the setting of the prediction model and was distributed via RedCap [23]. The questions were divided into 5 sections: general opinions, advantages and disadvantages, effects on patients, effects on colleagues, and opinions on implementation. The questionnaire consisted of 21 questions on a Likert scale and 1 open question.

Data Analysis

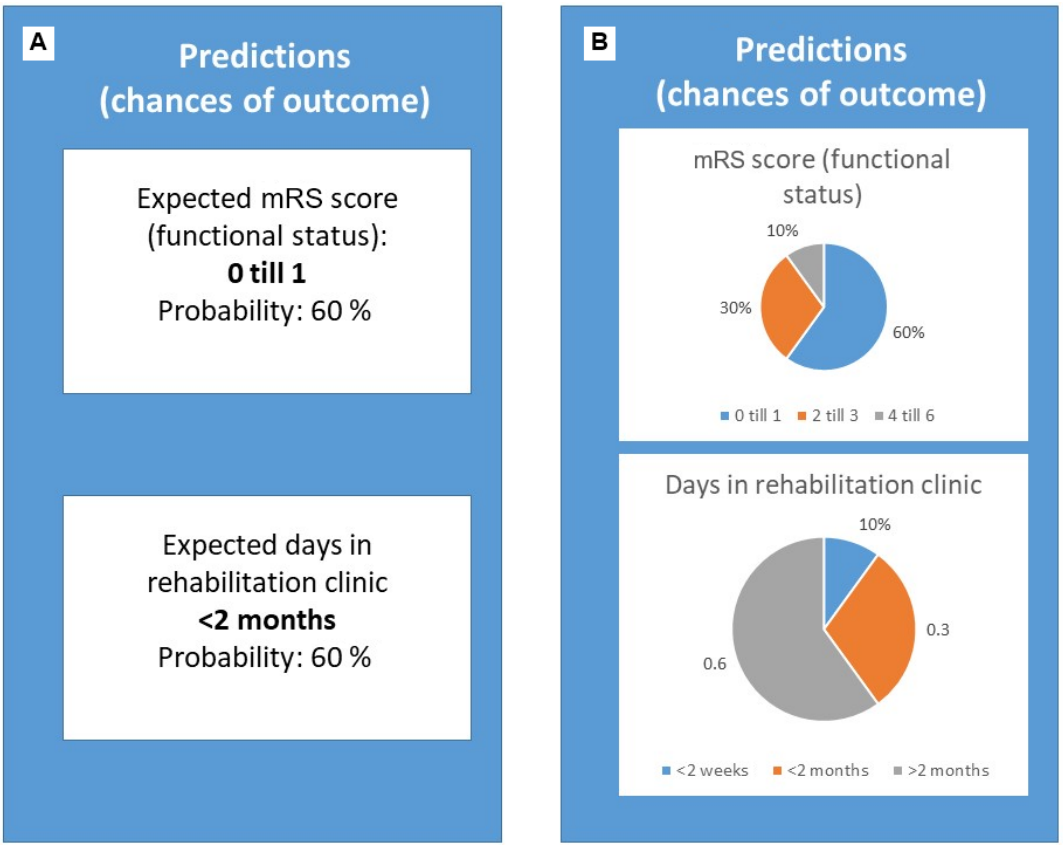
No statistical analyses were performed on the questionnaire, as the MIDI questionnaire is not a validated questionnaire and the sample size is small. Instead, the questions were meant to provide a formalized evaluation of the prediction model and its visualizations. Both the general score of the prediction model and the agreement have been reported.

Prediction Model and Visualizations

The model visualizations were based on the model described in a previous article [24]. This model aims to predict a 2-fold outcome: functional outcome (modified Rankin scale [mRS] score) after 3 months and length of stay in a rehabilitation center. Two feature sets were used as predictors: hospital data and data from geriatric rehabilitation clinics. The hospital data were based on registry data from the Dutch Acute Stroke Audit [25], such as medical history, treatment data, and admission data, and the other data included intake data, admission data, and discharge data.

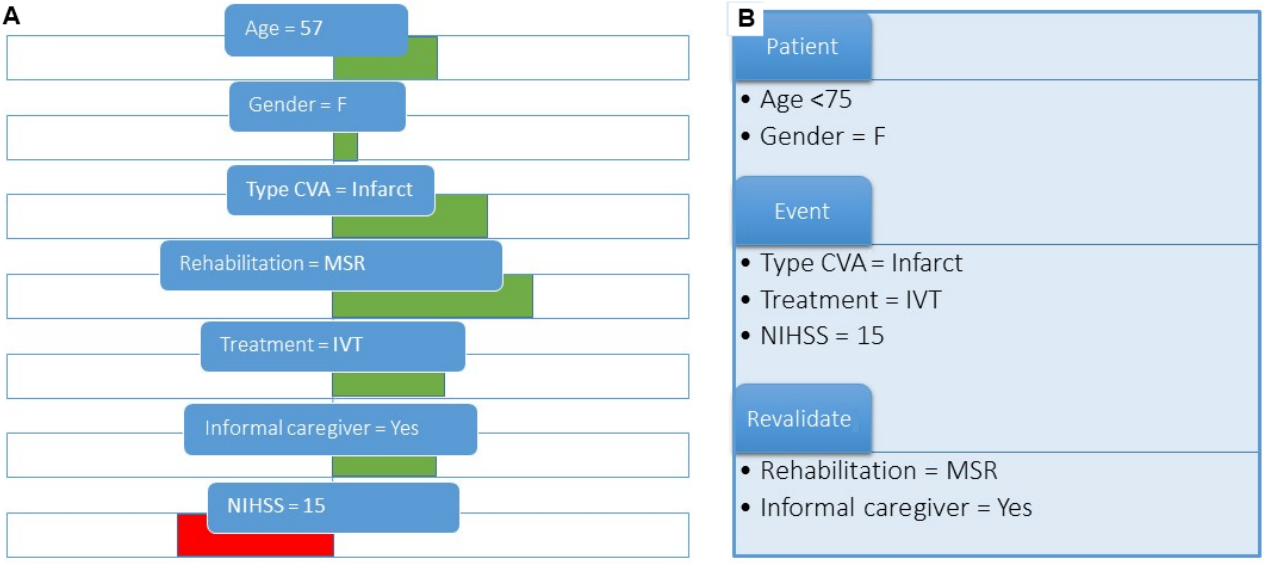
The design of the visualizations followed an iterative setup with the research team. Before the focus groups, the research team created a mock-up, with outcomes and visualizations for 2 fictional patients, as a functional model with real patient data was not yet available. Based on examples and previous literature about explainable AI in health care [26,27], a choice of 2 different visualization styles was made. One with a more simplified textual approach having clear and limited information, and one with a more detailed visualized approach having more information in the form of diagrams. We chose circle and bar charts for this style, as they are considered the most “approachable style” of visualization of numbers [19]. The visualizations of the outcomes in both styles are shown in Figure 2.

Figure 2. Different visualizations of outcomes (translated from Dutch). (A) Textual limited information; (B) Visual detailed information. mRS: modified Rankin scale.



Next to outcome visualizations, we also included explanations of the predicted outcomes by the models, which have been shown to be important for the adoption of AI in medicine [14]. We opted to show the model explainability through feature importance, where for the predicted outcome of the model for each patient, the contribution of each feature is shown. This can also be described as local interpretability, which has been further explained previously [19,23]. For example, it can show that the age of that specific patient contributed 20% to the decision of their personal outcome. This can be both positive or negative feature importance, indicating whether it made the predicted outcome more or less likely. The detailed visualization (Figure 3A) shows this in red and green bars. For the simplified version (Figure 3B), only the most important features for the outcome were mentioned (not the extent of their contribution).

Figure 3. Different visualizations of influencing factors (translated from Dutch). (A) Visual detailed information. The color and length of the bars indicate the size of the effect on the outcome. (B) Textual limited information. CVA: cerebrovascular accident; IVT: intravenous thrombolysis; MSR: medical specialistic rehabilitation; NIHSS: National Institutes of Health Stroke Scale.



Visualizations were developed for 2 fictional patients in 2 different visualization styles, resulting in 4 examples. The complete translated visualizations can be found in Multimedia Appendix 3. Based on the feedback from the focus groups, 1 visualization style was selected, and the visualizations were modified in accordance with the focus group feedback. These final visualizations were used for the questionnaire.

Ethical Considerations

The Medical Ethics committee of Utrecht (MEC-U) declared that this study is not subject to the Dutch Medical Research Involving Human Subjects Act (WMO) and issued a non-WMO statement under number W22.007, and subsequently, the board of directors of St. Antonius Hospital issued a statement of no objection (permission to perform) under number Z22.009. All participants provided informed consent.

Results

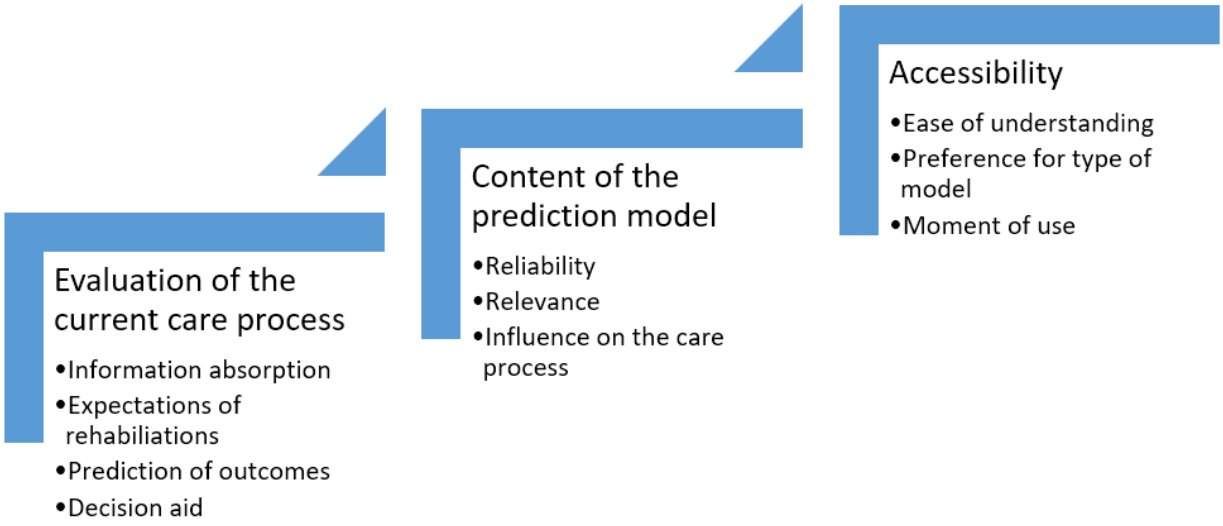
Focus Groups

The focus groups took place in April and May 2022 and lasted for 60-90 minutes with patients and caregivers and for about an hour with HCPs. We conducted 3 focus groups of each type (total 9 patients, 4 informal caregivers, and 8 HCPs). After these sessions, further focus groups were not deemed necessary, as saturation was reached.

A few minor changes were made to the interview guide after the first focus group with patients. We put less focus on the discharge interview and more on the questions about the care process in general for patients and the family interview for HCPs. For the focus groups with patients and informal caregivers, the transcripts were sent to these participants for a member check. None of the participants had any remarks on the transcripts.

The results of the focus groups have been organized by theme in prediction model development in 3 sections: evaluation of the current care process, content of the prediction model, and accessibility of the model (Figure 4).

Figure 4. The 10 themes in the focus groups divided into 3 categories.



Evaluation of the Current Care Process

Information Absorption

Some of the patients said that they either could not remember or did not understand the information given at the discharge interview. This could be due to the symptoms of their CVA. For example, a patient developed aphasia because of the CVA. As a result, she did not know what questions to ask during the discharge interview and did not manage to do so. However, most patients said this was due to information overload caused by discharge taking place shortly after the CVA, and the impressions and emotions during admission. On the other hand, most of the informal caregivers said that they received the information provided during the discharge interview and helped the patients with the information when necessary.

The patients and informal caregivers who could remember said that they were positive about the information provision, calling it clear, open, and informative, with a lot of room for questions. HCPs mentioned that they often found it hard to estimate whether patients and caregivers truly understood the given information, saying that patients always say they understand, even if they do not. They also mentioned language barriers as a common issue. Moreover, they were not sure whether the extra resources given, such as websites or information flyers, were read and understood by patients, as they did not discuss these with them.

Expectations of Rehabilitation

Most patients and informal caregivers had little idea of what to expect from rehabilitation and expressed that they felt thrown in at the deep end. One patient described her lack of expectations as follows:

> *I just let everything wash over me. That was all I could do.*

Only 1 patient said he knew what he could expect in his personal situation, even if daily tasks proved more difficult. Informal

caregivers expressed wishes for more information from the hospital about rehabilitation, with 1 informal caregiver expressing that he would have wanted a rehabilitation plan from the hospital:

> *[...] my mom is not the first one with a CVA, so there is a certain expectation and based on that […] protocols are made. It's my first time, and then there's nothing […] that's hard. Something has to be done about that. That's what I missed.*

Eventually, most patients were happy with the rehabilitation, except for a few bad experiences with HCPs. HCPs' experiences matched with the views of the patients and caregivers, and they stated that they do not have a good overview of rehabilitation and therefore cannot give patients an expectation of what will happen during rehabilitation. Some HCPs believed that this information should be given in rehabilitation clinics. They also mentioned that there could be a disparity in the expectations between patients and informal caregivers. In those situations, patients often felt like they were ready to be discharged home, while caregivers would still be very worried and push for inpatient rehabilitation.

Predictions of Outcomes

Almost all HCPs mentioned the difficulty of making individual predictions. They explained that this was due to patients, even similar patients, having varying recovery paths and because they did not have oversight over recovery during rehabilitation. Only the 2 most experienced neurologists felt more comfortable giving general predictions using age and the severity of the event. One of the junior doctors mentioned that they did not feel like they had enough knowledge and data to make such predictions.

> *I also don't know the exact scientific evidence from the top of my head […] for someone with a hemiparesis after 3 months.*

These statements were in agreement with the experiences of the patients. In the perception of the patients, if the physician did predict outcomes, it was not very precise. For example, they mentioned that the physician predicted that they would recover gradually in the first 3 months. Moreover, they said that no data in the form of numbers or percentages were used.

We asked all participants what they considered the most important outcomes to have a prediction on. Three outcomes were mainly mentioned: functional status in the medium or long term (multiple months or years), time of rehabilitation, and the chance of a new CVA. A new event was something that many patients were very scared about. They wanted to know how high the chance of recurrence was and wanted reassurance that it was as low as possible. HCPs mentioned it was not necessary to have a prediction of recurrence, as medications will always be given to minimize those chances.

Decision Aid

Many patients and informal caregivers felt like the physician took their opinions into account. According to the participants, it varied for each patient whether they had a choice in discharge location. Most participants said they did not have a choice in this decision. However, most participants felt that they could not make a better decision than a physician could, and they accepted the physician's opinion and professionalism. According to the participants, physicians have more knowledge about a CVA than patients and therefore know what is better for them.

> *You have to imagine; it was all new to us. So […] how can your opinion be different from what they recommend to you? It's almost impossible because you're in a situation like that so you can't judge it.*

Physicians mentioned shared decision-making. There is often not really any discussion on where the patient should be discharged.

Content of the Prediction Model

Reliability

Many patients were initially skeptical about a model that would predict outcomes. This was based on the physician's statement that it is not possible to make personal predictions. Several patients and informal caregivers said they did not believe that a physician could make accurate predictions and that consequences could only be seen after rehabilitation. Furthermore, some caregivers also found the prediction too absolute and programmed.

> *If it's such a well-founded prediction but we asked the neurologist about the whole thing, and he couldn't really give us any answers about predictions […]. Won't we get the wrong expectations from an overview like this?*

HCPs were also skeptical of the reliability of the model. They similarly argued that they see a lot of variation between patients and wondered if it is possible to make accurate predictions. In general, they believed that all included predictors would be usable. However, the HCPs questioned the reliability of the model due to the absence of certain important predictors, such as pre-existent functioning (Barthel score), detailed comorbidity, and the character of the patient. They acknowledged that the last aspect would be hard to measure. Moreover, if it is just a tool for information, the doctors argued that it is not that bad if the prediction is not 100% accurate**.**

> *[perfect performance] would be preferable, but you can never predict that […] maybe the important part is how to translate model performance to the patient*

One of the informal caregivers felt that there should be a disclaimer with the model because the outcome could never be 100% right.

Relevance

While most HCPs were generally enthusiastic about the proposed model, another point of skepticism was regarding relevance, as they were hesitant about how to explain the predicted outcomes to patients, especially regarding percentages. Patients confirmed these concerns. One of the patients said he did not like percentages because one can never say what 90%, for example, truly quantifies. Patients agreed that it is better to have a conversation and explain the information instead of simply showing the model. HCPs mentioned that while a general model for outcome prediction would be good, a more useful option would be if there were more different outcomes. For

example, where it would be more useful if we could split it into several types of complaints or symptoms (like aphasia or paralysis).

There was disagreement among patients regarding whether they wanted predictions at all. Some stated that they would be curious to know their predictions. Another participant stated that the information he received from his physician was sufficient and that he did not need a more specific prediction. Some of the caregivers said they would find it convenient to use a prediction model during a discharge conversation and that a prediction could supplement the information given by HCPs. Most HCPs were slightly worried about the risk of giving too much information, as it is currently already too much. However, some physicians also thought that a prediction could help clarify the situation, making the other information more digestible.

Influence on the Care Process

There were several points discussed concerning the possible influence of a prediction model on the care process. There was agreement among HCPs that it would be good for the patient as guidance if it is understandable for the patient. Junior doctors especially thought it would be good for HCPs as well for guidance in conversations with patients, reaching a better ground to make prognoses and overcoming unconscious biases about patients.

While most patients and informal caregivers expressed the positive aspects of guidance similar to HCPs, some patients worried that if the prediction is negative, it could be scary or it could demotivate the rehabilitation process. The patients stated that a lot of motivation is necessary for the best recovery.

> *And if I were [example patient] and 82 [years old] I don't know if I would be up for knowing […] that bad outcome*

However, most patients and informal caregivers did not want HCPs to hold back negative results from them. Moreover, doctors mentioned they would not be worried about this, as they are used to delivering bad news.

The main importance for HCPs was a way to inform patients and to offer clearer information for expectation management in the further rehabilitation process. As such, using this prediction tool would not have a direct result on the treatment or outcome of patients. There was the question of why this would be useful, and it was expressed that the goal of the tool should be clear; otherwise, it might not be used.

Accessibility

Ease of Understanding

The HCPs in general found the prediction model and the visualizations quite clear, and most thought that with explanation the model should be understandable for patients. However, several patients and informal caregivers found it hard to understand. Even after another explanation, they struggled to fully comprehend the meaning. One patient mentioned this might be due to one of the complications of his CVA being a lack of concentration and thus not having enough concentration to read. They struggled with the medical terminology, even after explanation. One patient said that it might be convenient for a physician but it is too hard for a patient. Most patients and caregivers understood the visualizations if they looked at them for a little longer but concurred that an explanation is needed from someone with a medical background or a more detailed but simple explanation is needed in additional text.

Preference for a Model Style

There were differences among patients regarding the preferences for the style of visualizations. Most of the participants preferred the prediction model with more graphics and less text because it was more visual and easier to understand. One participant stated that when a patient has symptoms like trouble with reading or concentration, it is easier to see the graphs.

> *I also think that for people who have […] more symptoms, who understand it a […] less easily […] the pictures are a little bit more informative.*

A few patients said the bar charts for feature importance did not have their preferences, as they contained too much information and the features could not be changed.

All HCPs preferred the more detailed visualization, and they argued that they contained more information and made the prediction more "alive." Interestingly, they thought patients would not have the same preferences. The HCPs also argued for integration into the electronic health record (EHR), as this would help with ease of use and adoption in the care process. Some HCPs were afraid that due to the high workload and the model not having a direct impact on the care process, the model might not be used if it is not easily accessible.

Moment of Use

Whether to use the model during the discharge interview or during the family interview, which happens a few days into admission with informal caregivers, was a point of discussion among HCPs. It could be beneficial for both. The discharge interview can be very rushed, and during the family interview, there might be more time to go over the predictions. However, this would be earlier in the process, which means that less information might be included in the predictions. Moreover, they agreed that having some form of access afterward would be good. Different options were mentioned: at the rehabilitation center, with the general practitioner, or at the 3-month follow-up.

The patients generally agreed with this view. One patient stated that it would be good to see the prediction when you feel like it, but it would have been too much during the discharge interview.

> *It is useful but […] too extensive [for] a normal person […] does not understand a thing. [...] Then it is explained by someone with a medical background […] who can perhaps explain it better. […] Maybe after a few months go back and look at it again […]. What are the expectations now, what are the expectations later.*

Questionnaire Survey

Based on the results of the focus groups, we extracted a set of recommendations. These recommendations were either immediately included in the visualizations of the prediction

model or served as general recommendations for further development or for the context of the implementation in current care. The recommendations are summarized in Table 2.

Table 2. Recommendations from the focus groups.

Category	Theme	Recommendation	Incorporation
Evaluation of current care	Information absorption	Recognize that the information might not be understood at the discharge interview	Implementation
Evaluation of current care	Expectations of rehabilitation	Create a better overview of rehabilitation for patients, informal caregivers, and health care professionals	Context of care
Evaluation of current care	Prediction of outcomes	Highlight that the model meets the needs of patients, informal caregivers, and health care professionals regarding providing a prediction of the outcome	Immediately
Evaluation of current care	Decision aid	Make sure the patient's opinion is heard in the decision process, for example, through shared decision-making	Context of care
Accessibility	Ease of understanding	The target audience of visualizations should be health care professionals and not patients	Immediately
Accessibility	Moment of use	Do not limit the use of prediction to the discharge interview	Implementation
Accessibility	Model preference	Visualize outcome and feature importance (visualization version b)	Immediately
Content of the model	Reliability	Include the general predictive performance of the model, for example, according to the margin of error	Future research
Content of the model	Relevance	mRS[a] score in 3 categories, including the possible functional score, measured during hospital stay	Future research
Content of the model	Influence on the care process	Clarify the goal of the model (informing patients) to improve adherence	Immediately

[a]mRS: modified Rankin scale.

The visualizations were adapted according to the recommendations. We selected the more "visualized" versions (see Figures 2B and 3A). Moreover, we focused on HCPs as our target audience and therefore removed explanations of well-known abbreviations and concepts, such as the National Institutes of Health Stroke Scale (NIHSS) score. We also specified that the moment of use could be the discharge or family interview and that the goal of the prediction model was to inform patients. Moreover, we included more detailed information on feature importance, based on our underlying model. The adapted visualizations can be found in Multimedia Appendix 3.

The questionnaire was sent to 16 relevant HCPs in St. Antonius Hospital, and we received 9 responses. The answers to the closed questions are summarized in Table 3. For each question, we have reported scores. Moreover, we have presented the average score and SD for each section. Negatively phrased questions (questions 4, 8, and 9) were scored in reverse.

Table 3. Questionnaire results reported on a Likert scale (0-5).

Question[a]	Score, mean (SD)
General statements	
1. The prediction model matches well with my usual way of working.	4.0 (0.5)
2. The prediction model is based on factually correct information.	3.7 (0.5)
3. The prediction model offers all the information and materials necessary to work with it properly.	3.6 (0.5)
4. The prediction model is too complicated for me to use.	4.1 (0.8)
5. I expect that the effects of the use of the prediction model will be clearly visible.	2.9 (0.4)
6. I think the prediction model is suitable for my patients.	3.8 (1.0)
Advantages and disadvantages	
7. The use of the prediction model gives me the opportunity to inform my patients better.	3.8 (1.0)
8. The use of the prediction model during discharge or family interviews will cost me more time than usual.	3.2 (0.5)
9. I find it too complicated to use the prediction model during discharge or family interviews.	3.9 (0.8)
Prediction model and patients	
11. I expect that with the prediction model, my patients will be better informed about the expected functional status after 3 months and the duration of rehabilitation.	3.4 (0.5)
12. I think it is important to use the prediction model to better inform my patients about the expected outcomes: the functional status after 3 months, and the duration of rehabilitation.	4.0 (0.5)
13. I consider it part of my job to use the prediction model.	3.6 (0.7)
14. Patients will generally be satisfied if I use this prediction model.	3.8 (0.4)
15. Patients will generally be cooperative if I use this prediction model.	3.7 (0.5)
Prediction model and colleagues	
16. I can count on sufficient help from my colleagues when necessary while using the prediction model.	3.5 (0.5)
17. How many doctors who conduct discharge interviews will actually use this prediction model?	3.8 (0.7)
18. To what extent will (your fellow) neurologists expect you to use the prediction model?	3.2 (0.4)
19. When it comes to working with the prediction model, how much do you care about the opinion of (your fellow) neurologists?	3.7 (0.5)
Application of the prediction model	
20. If you wanted to, do you think you would be able to apply the prediction model during the discharge or family discussions?	3.9 (0.3)
21. I have sufficient knowledge to be able to use the prediction model.	3.8 (0.8)
22. To what extent are you aware of the content of the prediction model?	3.1 (1.1)

[a]Question 10 was an open question (What are the other advantages or disadvantages of this prediction model?) and is not included in the table.

We considered statements with a score of ≥3.5 and an SD of <1.0 to be sufficient to positive, with agreement among HCPs. Most statements were received as such, and we considered the general feedback to be moderately positive. Moreover, in all categories of statements, we noted that most statements were at least moderately positive with agreement, highlighting that there is no topic where the prediction model consistently underperforms. The high scores for "*The prediction model matches well with my usual way of working*" (score 4.0), "*I think it is important to use the prediction model to better inform my patients about the expected outcomes: the functional status after 3 months, and the duration of rehabilitation*" (score 4.0), and "*If you wanted to, do you think you would be able to apply the prediction model during the discharge or family discussions?*" (score 3.9) indicate a high willingness to implement the prediction model.

For some questions/statements, the HCPs were not positive on average (score of <3.5) or were not in agreement (SD ≥1.0). The statement "*I expect that with the prediction model, my patients will be better informed about the expected functional status after 3 months and the duration of rehabilitation*" had a score of 3.4. This was just under our set threshold. Interestingly, the statement "*I think it is important to use the prediction model to better inform my patients about the expected outcomes: the functional status after 3 months, and the duration of rehabilitation*" had a score of 4.0. It shows that the HCPs, even when not sure about the positive outcome of use, still found it important to use the prediction model. The same concern was highlighted in 2 statements that had high disagreement among the HCPs: "*I think the prediction model is suitable for my patients*" (score 3.8, SD 1.0) and "*The use of the prediction model gives me the opportunity to inform my patients better*" (score 3.8, SD 1.0). This shows consistency with the focus

groups indicating that there seems to be variety in the extent to which HCPs are positive about the effects of the prediction model.

In the other statements that were not positive on average or had high disagreement, we also noted similar findings that came up in the focus groups as well. The statement "*Iexpect that the effects of the use of the prediction model will be clearly visible*" had the lowest score of 2.9. This is not completely unexpected as the model's primary function is to inform and not provide decision support. Moreover, the issue of high workload was reflected in the statement "*The use of the prediction model during discharge or family interviews will cost me more time than usual*" (score 3.2). Lastly, the statement "*To what extent are you aware of the content of the prediction model?*" had both a low score and high disagreement (score 3.1, SD 1.1), indicating that more detailed explanation and training are necessary to understand the model.

The statement "*To what extent will (your fellow) neurologists expect you to use the prediction model?*" had a low score of 3.2. This shows that many HCPs would consider the use of such a prediction model to be more elective. Next to the scaled questions, we also asked 1 open question "*what are the other advantages or disadvantages of this prediction model?*" A big point of contention was whether the prediction model would be accurate enough, as certain factors like recovery during the time in the hospital and mental resilience were not included. A positive aspect mentioned a few times was that it could also be used to preinform the rehabilitation clinics for helping with planning.

Discussion

Principal Findings

This study shows the added value of a prediction model for long-term outcomes of rehabilitation after stroke in CVA care under certain conditions. The need for a prediction model was consistently shown in focus groups and a questionnaire survey, as evidenced by the general approval of participants. Several recommendations can be made, most importantly on the following 5 conditions: primary target, timing of discussing the predictions, different focuses of HCPs and patients, goal of the prediction model, and visualization of the model.

First, the primary target audience of the prediction model should be HCPs. While the information provided is of interest to both patients and informal caregivers, not all patients can fully understand the outcomes of the prediction model. This resembles the findings of other studies that investigated different digital health tools for stroke patients, which reported complaints of difficulty [28] or disinterest among some patients [29]. Our study specifically found that patients prefer HCPs to provide information and guide them through the visualizations of the model.

Second, the timing for using predictions should also be considered. As patients mentioned, it could be more useful when targeted toward patients at different points in the care process. Here, it is interesting to see if the model can be simplified, as was requested by interested patients, and how that should be offered. Another study on digital health tools for patients with brain injuries [30] highlighted the importance of user testing among these patients. This would be a good way to ensure that the simplified model fits the patient's needs and understanding.

Third, for HCPs, the reliability and relevance of the model are key. HCPs focus largely on the reliability and relevance of the model. It is essential to not only provide good reliability and relevance but also clearly communicate the aspects to HCPs. Reliability is essential to validate the performance of the model. Therefore, changes in outcome reporting were considered to be more in line with the current care [31]: divide the mRS score into 3 categories and include predictive factors. By including the general predictive performance of the model (eg, according to the margin of error), the reliability can be better communicated to HCPs. These 2 points of reliability and relevance are very much in agreement with the guidelines for the implementation of AI in health care [32,33]. While the performance and generalizability of the model were not the focus of this study, they should be properly validated before model implementation. We found that patients focused more on the impact on rehabilitation, the emotional aspects, and understandability.

Fourth, HCPs and patients should understand the goal of the prediction model. This understanding can improve adherence and the willingness of HCPs to cooperate, which are essential for the adoption of the model in practice [32]. This matched our results, as this was one of the main points of the HCPs in both focus groups and the questionnaire survey. We aimed to clarify the goal of the prediction model after the focus groups. However, the questionnaires showed that the goal of the prediction model was still unclear to some HCPs, even after a more detailed explanation in the introduction of the questionnaire. This could be mitigated by properly informing HCPs, for example, by providing in-person training or a detailed demonstration of the model. It is also essential to highlight the goal of the explainability aspect of the model. The visualizations showed to what extent certain factors contributed to the model's decision, but that did not mean changing those factors would necessarily lead to a different outcome for the patient. For example, while the model might assign high importance to a certain treatment in the prediction of negative outcomes, it does not mean that not performing this treatment would lead to a better outcome.

Finally, all participant groups showed a preference for a more visual model. This matches the general consensus in visualization science, which argues for presenting numerical data in simple visualizations, such as bar or pie charts [15]. Our results showed that this is even more relevant for patients who have experienced a stroke. However, the visualizations that we used relied on red and green bars to highlight the importance of features, which can be an issue for colorblind people [34]. While none of our participants reported issues related to colorblindness, it should be adapted to a more inclusive visualization.

Strengths and Limitations

One of the main strengths of this study was the involvement of different stakeholder groups, including patients, informal

caregivers, and HCPs, with different roles. Moreover, the double setup with separate focus groups and questionnaires allowed for a check of agreement between the results of the different methods. One limitation to consider is our use of convenience sampling, which could have skewed the results. People with very severe stroke or with trouble reading or understanding were not represented. For HCPs, we mainly included people who were already open to research and new developments. The interviews with patients and informal caregivers were performed several months after the CVA, so their actual responses at the time of discharge might have been different. Moreover, we used mock-ups for the visualizations and not "real" predictions based on actual patient data. Although the visualizations were based on our existing models and data as much as possible, this was not a personal prediction. This made the visualizations more difficult to understand than expected, and the focus groups might have been more efficient with a functional model based on the actual patient data.

For the questionnaires, we had a relatively small sample size, and therefore, we did not perform any statistical analyses. As there was high agreement among the HCPs and the majority (9/16) of invited HCPs responded, it is likely they are a representative sample of the HCP group in the hospital, but we cannot know this for sure. Moreover, this group did not include all HCPs involved in stroke care. It is possible that physical therapists or general practitioners would have a different outlook on the prediction models, as they might be more involved in the rehabilitation of patients or involved at a different point in rehabilitation care. Lastly, this study was limited to a single hospital in the Netherlands. Differences between hospitals and especially between medical practices among countries can impact the generalizability of our results. Combining these limitations, we would recommend a quantitative follow-up study based on our questionnaire. Preferably, this should be a multicenter and international study to confirm the generalizability of our results outside the confines of our hospital.

Implications and Future Directions

From this study, we can extract lessons to be considered when implementing AI in medicine in general. An important aspect is that the need for information provision is highly variable between groups as well as individuals. The results of the prediction model can be highly personal, and similarly, the needs of a patient and the setting and specific illness of the patient can be personal. It is essential to clearly define the purpose of the tool, as there are not only different levels of understanding but also different priorities. For example, patients do not always appreciate detailed information, especially if the information is not actionable but only informative. It is also essential to consider a model that can have multiple purposes. A model, such as ours, could also be used for advanced resource planning in rehabilitation clinics by estimating the future demand on the clinics and evaluating the effectiveness of the rehabilitation.

An important point is to investigate how to include patients who struggle with technology due to language, socioeconomic, or medical issues. For example, this study excluded non-Dutch speakers, but HCPs mentioned that language issues could further complicate explaining the expectations of recovery. This difficulty in communication should not be increased by a new technological tool or prediction model. Relying more on technology can lead to more inequality in care [35], especially for vulnerable patients, such as stroke patients [32]. One effort that can prevent such issues is the development of a comprehensive framework for AI implementation in health care and the creation of a roadmap for AI implementation. Such frameworks are being developed [31,33], but none of them are exhaustive, and having resources that detail considerations for AI application and evaluation are still necessary to develop AI with meaningful medical applications [33].

Conclusions

The findings of this study show that participants are generally positive toward a prediction model for long-term outcomes of rehabilitation after stroke in CVA care under certain conditions, with a general preference for a more visualized prediction model. The prediction model should be geared toward HCPs, as they can provide the context necessary for patients and their informal caregivers. For HCPs, good reliability and relevance of the prediction model are essential for its proper integration. We recommend a quantitative follow-up study to confirm these results in multicenter and international settings.

Acknowledgments

This study is part of the project Enabling Personalized Interventions (EPI) with project number 628.011.028 of the research program Commit2Data-Data2Person, which is partly financed by the Dutch Research Council (NWO).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Guide for focus groups with patients (translated from Dutch).
[DOCX File , 16 KB - humanfactors_v12i1e56521_app1.docx]

Multimedia Appendix 2
Guide for focus groups with health care professionals (translated from Dutch).

[DOCX File , 15 KB - humanfactors_v12i1e56521_app2.docx]

Multimedia Appendix 3
Finalized visualizations (translated from Dutch).
[PNG File , 537 KB - humanfactors_v12i1e56521_app3.png]

References

1. Pan B, Jin X, Jun L, Qiu S, Zheng Q, Pan M. The relationship between smoking and stroke: A meta-analysis. Medicine (Baltimore) 2019 Mar;98(12):e14872 [FREE Full text] [doi: 10.1097/MD.0000000000014872] [Medline: 30896633]
2. Hankey GJ. Stroke. Lancet 2017 Feb 11;389(10069):641-654. [doi: 10.1016/S0140-6736(16)30962-X] [Medline: 27637676]
3. Geerars M, Wondergem R, Pisters MF. Decision-making on referral to primary care physiotherapy after inpatient stroke rehabilitation. J Stroke Cerebrovasc Dis 2021 May;30(5):105667 [FREE Full text] [doi: 10.1016/j.jstrokecerebrovasdis.2021.105667] [Medline: 33631474]
4. Armstrong MJ. Shared decision-making in stroke: an evolving approach to improved patient care. Stroke Vasc Neurol 2017 Jun;2(2):84-87 [FREE Full text] [doi: 10.1136/svn-2017-000081] [Medline: 28959495]
5. Krishnan S, Hay CC, Pappadis MR, Deutsch A, Reistetter TA. Stroke survivors' perspectives on post-acute rehabilitation options, goals, satisfaction, and transition to home. J Neurol Phys Ther 2019 Jul;43(3):160-167 [FREE Full text] [doi: 10.1097/NPT.0000000000000281] [Medline: 31205230]
6. Prick JCM, Zonjee VJ, van Schaik SM, Dahmen R, Garvelink MM, Brouwers PJAM, Santeon VBHC stroke group. Experiences with information provision and preferences for decision making of patients with acute stroke. Patient Educ Couns 2022 May;105(5):1123-1129 [FREE Full text] [doi: 10.1016/j.pec.2021.08.015] [Medline: 34462248]
7. Smith M, Ackerley S, Barber P, Byblow W, Stinear C. PREP2 algorithm predictions are correct at 2 years poststroke for most patients. Neurorehabil Neural Repair 2019 Aug;33(8):635-642. [doi: 10.1177/1545968319860481] [Medline: 31268414]
8. Smith M, Barber AP, Scrivener BJ, Stinear CM. The TWIST tool predicts when patients will recover independent walking after stroke: an observational study. Neurorehabil Neural Repair 2022 Jul;36(7):461-471. [doi: 10.1177/15459683221085287] [Medline: 35586876]
9. Harari Y, O'Brien MK, Lieber RL, Jayaraman A. Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. J Neuroeng Rehabil 2020 Jun 10;17(1):71 [FREE Full text] [doi: 10.1186/s12984-020-00704-3] [Medline: 32522242]
10. O'Brien M, Lanotte F, Khazanchi R, Shin S, Lieber R, Ghaffari R, et al. Early prediction of poststroke rehabilitation outcomes using wearable sensors. Phys Ther 2024 Feb 01;104(2):pzad183 [FREE Full text] [doi: 10.1093/ptj/pzad183] [Medline: 38169444]
11. Bonkhoff AK, Grefkes C. Precision medicine in stroke: towards personalized outcome predictions using artificial intelligence. Brain 2022 Apr 18;145(2):457-475 [FREE Full text] [doi: 10.1093/brain/awab439] [Medline: 34918041]
12. Wang W, Kiik M, Peek N, Curcin V, Marshall IJ, Rudd AG, et al. A systematic review of machine learning models for predicting outcomes of stroke with structured data. PLoS One 2020;15(6):e0234722 [FREE Full text] [doi: 10.1371/journal.pone.0234722] [Medline: 32530947]
13. Campagnini S, Arienti C, Patrini M, Liuzzi P, Mannini A, Carrozza MC. Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review. J Neuroeng Rehabil 2022 Jun 03;19(1):54 [FREE Full text] [doi: 10.1186/s12984-022-01032-4] [Medline: 35659246]
14. Herzog L, Kook L, Hamann J, Globas C, Heldner MR, Seiffge D, et al. Deep learning versus neurologists: functional outcome prediction in LVO stroke patients undergoing mechanical thrombectomy. Stroke 2023 Jul;54(7):1761-1769. [doi: 10.1161/STROKEAHA.123.042496] [Medline: 37313740]
15. Borsos B, Allaart CG, van Halteren A. Predicting stroke outcome: A case for multimodal deep learning methods with tabular and CT Perfusion data. Artif Intell Med 2024 Jan;147:102719 [FREE Full text] [doi: 10.1016/j.artmed.2023.102719] [Medline: 38184355]
16. Combi C, Amico B, Bellazzi R, Holzinger A, Moore JH, Zitnik M, et al. A manifesto on explainability for artificial intelligence in medicine. Artif Intell Med 2022 Nov;133:102423 [FREE Full text] [doi: 10.1016/j.artmed.2022.102423] [Medline: 36328669]
17. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310. [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
18. Vasey B, Nagendran M, Campbell B, Clifton DA, Collins GS, Denaxas S, DECIDE-AI expert group. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. Nat Med 2022 May;28(5):924-933. [doi: 10.1038/s41591-022-01772-9] [Medline: 35585198]
19. Trevena LJ, Bonner C, Okan Y, Peters E, Gaissmaier W, Han PKJ, et al. Current challenges when using numbers in patient decision aids: advanced concepts. Med Decis Making 2021 Oct;41(7):834-847 [FREE Full text] [doi: 10.1177/0272989X21996342] [Medline: 33660535]

20. ATLAS.ti Scientific Software Development GmbH. Atlas.ti. URL: https://atlasti.com/ [accessed 2024-12-18]
21. Green J, Thorogood N. Qualitative Methods for Health Research. Thousand Oaks, CA: SAGE Publications; 2018.
22. Fleuren MAH, Paulussen TGWM, Van Dommelen P, Van Buuren S. Towards a measurement instrument for determinants of innovations. Int J Qual Health Care 2014 Oct;26(5):501-510 [FREE Full text] [doi: 10.1093/intqhc/mzu060] [Medline: 24951511]
23. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
24. Kassem J, Allaart C, Amiri S, Kebede M, Müller T, Turner R, et al. Building a Digital Health Twin for Personalized Intervention: The EPI Project. In: Commit2Data. Open Access Series in Informatics (OASIcs), Volume 124. Wadern, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik; 2024.
25. Kuhrij LS, Wouters MW, van den Berg-Vos RM, de Leeuw F, Nederkoorn PJ. The Dutch Acute Stroke Audit: Benchmarking acute stroke care in the Netherlands. Eur Stroke J 2018 Dec;3(4):361-368 [FREE Full text] [doi: 10.1177/2396987318787695] [Medline: 31236484]
26. Duell J. A Comparative Approach to Explainable Artificial Intelligence Methods in Application to High-Dimensional Electronic Health Records: Examining the Usability of XAI. arXiv. 2021. URL: https://arxiv.org/abs/2103.04951 [accessed 2024-12-18]
27. Ahmad M, Eckert C, Teredesai A. Interpretable Machine Learning in Healthcare. In: BCB '18: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2018 Presented at: ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; August 29, 2018 to September 1, 2018; Washington, DC, USA. [doi: 10.1145/3233547.3233667]
28. Epalte K, Tomsone S, Vētra A, Bērziņa G. Patient experience using digital therapy "Vigo" for stroke patient recovery: a qualitative descriptive study. Disabil Rehabil Assist Technol 2023 Feb;18(2):175-184. [doi: 10.1080/17483107.2020.1839794] [Medline: 33155507]
29. Bally ELS, Cheng D, van Grieken A, Ferri Sanz M, Zanutto O, Carroll A, et al. Patients' perspectives regarding digital health technology to support self-management and improve integrated stroke care: qualitative interview study. J Med Internet Res 2023 Apr 04;25:e42556 [FREE Full text] [doi: 10.2196/42556] [Medline: 37014677]
30. Miao M, Morrow R, Salomon A, Mcculloch B, Evain J, Wright MR, et al. Digital health implementation strategies coproduced with adults with acquired brain injury, their close others, and clinicians: mixed methods study with collaborative autoethnography and network analysis. J Med Internet Res 2023 Sep 19;25:e46396 [FREE Full text] [doi: 10.2196/46396] [Medline: 37725413]
31. van de Sande D, Van Genderen ME, Smit JM, Huiskens J, Visser JJ, Veen RER, et al. Developing, implementing and governing artificial intelligence in medicine: a step-by-step approach to prevent an artificial intelligence winter. BMJ Health Care Inform 2022 Feb;29(1):e100495 [FREE Full text] [doi: 10.1136/bmjhci-2021-100495] [Medline: 35185012]
32. Verma A, Towfighi A, Brown A, Abhat A, Casillas A. Moving towards equity with digital health innovations for stroke care. Stroke 2022 Mar;53(3):689-697 [FREE Full text] [doi: 10.1161/STROKEAHA.121.035307] [Medline: 35124973]
33. de Hond AAH, Leeuwenberg AM, Hooft L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. NPJ Digit Med 2022 Jan 10;5(1):2 [FREE Full text] [doi: 10.1038/s41746-021-00549-7] [Medline: 35013569]
34. Katsnelson A. Colour me better: fixing figures for colour blindness. Nature 2021 Oct;598(7879):224-225. [doi: 10.1038/d41586-021-02696-z] [Medline: 34608306]
35. Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L. Inequities in health care services caused by the adoption of digital health technologies: scoping review. J Med Internet Res 2022 Mar 21;24(3):e34144 [FREE Full text] [doi: 10.2196/34144] [Medline: 35311682]

Abbreviations

AI: artificial intelligence
CVA: cerebrovascular accident
HCP: health care professional
MIDI: Measurement Instrument for Determinants of Innovations
mRS: modified Rankin scale

Edited by A Choudhury; submitted 22.01.24; peer-reviewed by A Hassan, T Abd El-Hafeez, R Gore, F Lanotte; comments to author 16.09.24; revised version received 29.10.24; accepted 20.11.24; published 22.01.25.

Please cite as:
Allaart CG, van Houwelingen S, Hilkens PHE, van Halteren A, Biesma DH, Dijksman L, van der Nat PB
The Significance of a Cerebrovascular Accident Outcome Prediction Model for Patients, Family Members, and Health Care Professionals: Qualitative Evaluation Study
JMIR Hum Factors 2025;12:e56521
URL: https://humanfactors.jmir.org/2025/1/e56521
doi:10.2196/56521
PMID:

Original Paper

Exploring the Users' Perspective of the Nationwide Self-Exclusion Service for Gambling Disorder, "Spelpaus": Qualitative Interview Study

Johanna Tjernberg[1], RN; Sara Helgesson[1], RN; Anders Håkansson[1], PhD; Helena Hansson[1], PhD

Department of Clinical Sciences Lund, Faculty of Medicine, Lund University, Lund, Sweden

Corresponding Author:
Anders Håkansson, PhD
Department of Clinical Sciences Lund
Faculty of Medicine
Lund University
Psychiatry, Baravägen 1
Lund, 221 00
Sweden
Phone: 46 703135677
Email: anders_c.hakansson@med.lu.se

Abstract

Background: Problem gambling and gambling disorder cause severe social, psychiatric, and financial consequences, and voluntary self-exclusion is a common harm reduction tool used by individuals with gambling problems.

Objective: The aim of this study was to explore users' experience of a novel nationwide, multioperator gambling self-exclusion service, "Spelpaus," in Sweden and to inform stakeholders and policy makers in order to improve harm reduction tools against gambling problems.

Methods: Semistructured interviews were conducted with 15 individuals who reported self-perceived gambling problems and who had experience of having used the self-exclusion service Spelpaus in Sweden. Interviews were transcribed and analyzed through qualitative content analysis.

Results: We identified 3 categories and 8 subcategories. The categories were (1) reasons for the decision to self-exclude, (2) positive experiences, and (3) suggestions for improvement. The subcategories identified a number of reasons for self-exclusion, such as financial reasons and family reasons, and positive experiences described as a relief from gambling; in addition, important suggestions for improvement were cited, such as a more gradual return to gambling post–self-exclusion, better ways to address loopholes in the system, and transfer from self-exclusion to treatment.

Conclusions: Voluntary self-exclusion from gambling, using a nationwide multioperator service, remains an appreciated harm-reducing tool. However, transfer from self-exclusion to treatment should be facilitated by policy making, and loopholes allowing for breaching of the self-exclusion need to be counteracted.

(JMIR Hum Factors 2025;12:e66045) doi:10.2196/66045

KEYWORDS

gambling disorder; gambling addiction; behavioral addiction; harm reduction; self-exclusion; voluntary self-exclusion; Spelpaus; lived experience; human factors; usability; qualitative study

Introduction

Background

Gambling disorder is a behavioral addictive disorder known to cause severe financial, social, and mental health consequences to affected individuals [1-3]. Among the harm-reducing or harm-preventive strategies used against gambling-related harm, voluntary self-exclusion is a well-established responsible gambling measure aiming to prevent and reduce the damage that can be caused by gambling [4].

Through self-exclusion, gamblers can voluntarily suspend themselves from gambling. However, 1 well-known limitation of self-exclusion services may be the possibility to gamble with another operator than the one the gambler has been self-excluded

from [5]. Gamblers are therefore usually required to self-exclude from multiple different operators. Consequently, in recent years, to further prevent gamblers from breaching their self-exclusion, it has become more common to offer gamblers the possibility to self-exclude from multiple operators at the same time.

In January 2019, the legislation for the Swedish gambling market changed, and monopoly was replaced by a license-regulated market [6]. This regulation meant that gambling operators, including overseas operators, needed to apply for a Swedish gambling license to operate on the Swedish market. In connection to the new regulations, the nationwide self-exclusion service "Spelpaus" was launched. Gambling operators with a Swedish license must adhere to this self-exclusion service, along with a series of other measures for responsible gambling. Spelpaus offers the possibility to self-exclude from all gambling operators with a Swedish license for 1, 3, 6, or 12 months. Self-exclusion cannot be canceled prematurely, and if the 12-month option is chosen, the self-exclusion continues beyond the 12 months, unless the individual actively chooses to cancel it after that period. Self-exclusion via Spelpaus also prohibits gambling operators from sending direct advertisements to self-excluded customers.

Spelpaus is in many ways a unique harm-reducing tool since it is nationwide, covers all licensed gambling operators (both online and land based), and can be accessed separately from the gambling operators. Thus, it theoretically represents a novel harm reduction strategy over and above other responsible gambling measures. Indeed, theoretically, this system may have a larger coverage and fewer loopholes (ie, fewer practical possibilities to turn into other gambling modalities and gamble despite self-exclusion) compared to a self-exclusion system involving 1 or few gambling operators. Around 100 gambling companies have a Swedish gambling license, with most of them operating exclusively online.

The prevalence of gambling problems in the Swedish population is estimated to be approximately 0.5% on a diagnostic level and a total of 1%-2% with at least moderate-risk gambling according to the Problem Gambling Severity Index [7,8]. Formal treatment seeking, or treatment provision, is low; less than 1000 people annually receive a gambling disorder diagnosis in the whole of Sweden, and even when taking other care providers into considerations, it can be assumed that only a small minority of people with a gambling problem on a diagnostic level receive formal treatment [9]. In contrast, low formal treatment seeking is often described in gambling disorder; however, a more informal but still active help-seeking behavior appears to be common, and voluntary self-exclusion from gambling can be part of this [10].

In Sweden, around 110,000 inhabitants are enrolled in Spelpaus [11], a number that has steadily increased since the launch in 2019 [12]. Thus, we believe that an increasing number of people with problematic gambling are attempting to intervene against their problem through self-exclusion, although we must consider that people without gambling problems also choose to self-exclude.

The steady increase in registrations in Spelpaus indicates that it is perceived as a helpful tool for many people. However, there are also limitations in the self-exclusion service, making it possible to breach the self-exclusion. Such a limitation is that Spelpaus only covers operators with a Swedish license, therefore making it possible for gamblers to gamble with overseas operators outside the Swedish license gambling market. In a 2020 web survey among gamblers, 38% of gamblers stated that they had gambled with unlicensed operators, while being self-excluded through Spelpaus [13], and in a later survey in 2022, the corresponding figure was 49% [14]. In another survey, conducted by the Swedish Gambling Authority, 25% gamblers stated that they had gambled with unlicensed operators during their self-exclusion [15]. In addition, continued gambling despite self-exclusion remains part of the clinical picture in clinical gambling disorder treatment in this setting [16].

Previous international experience of responsible gambling tools has demonstrated not only the feasibility of self-exclusion services and their use by individuals with gambling problems [4,17,18] but also that users experience limitations of these services [10,18]. Altogether, previous research calls for more in-depth examinations of user experience with self-exclusion services and suggestions for improvement of their effectiveness. In particular, this is relevant for the present type of self-exclusion service, which theoretically has been designed to overcome well-known limitations of older self-excluson systems; this novel Swedish service (1) is government based and independent of gambling operators, such that an individual does not need to enter a gambling site in order to self-exclude; (2) covers all licensed operators in the country and therefore prevents the possibility to change between licensed operators during an ongoing self-exclusion period; and (3) involves both land-based and online-based operators.

Thus, altogether, the relatively recent introduction of a unique, large-scale, multioperator self-exclusion service in Sweden has attracted a large number of users, but it is hindered by the risk of breaching one's self-exclusion to an extent that is clearly of relevance in the treatment setting. Therefore, it is of great value to study which effects of the Spelpaus service are perceived as favorable and challenging by its users and to also study this in more detail than in previous quantitative survey studies.

Aim

Based on the aforementioned research gap, the aim of this study was to enhance comprehension of the user experience of the Spelpaus service within the gambling community and to acquire a more profound insight into the reasons for use of this harm reduction tool and its advantages and challenges. Thereby, the overarching goal of the study was to inform policy makers about potential drawbacks and the potential for improvement of the harm reduction tool.

Methods

Study Design

This research was exploratory in nature. A semistructured interview guide was developed to ensure the exploration of diverse perspectives and experiences related to Spelpaus. Notably, this study represents the first attempt to comprehensively examine user experiences of Spelpaus, thus

justifying the use of qualitative content analysis for its methodological flexibility. The protocol of this study has been published previously [19].

An initial interview guide was developed based on the study's aim and research questions. A test interview with a peer support worker who had experience with self-exclusion services was conducted, leading to refinements of some questions to be more open ended. Further details of the interview questions are provided in the study protocol [19].

Context

The study was conducted in Sweden, a country with approximately 10.4 million inhabitants, where gambling is prevalent, with around 56% of individuals aged from 16 to 84 years having engaged in gambling activities in the past year. Moreover, 18% gamble at least once per month. The prevalence of gambling is higher among men aged 45-84 years, and online gambling has increased markedly in popularity in Sweden over the past decade [20]. Problem gambling in the present setting is predominated by online casino gambling and sports betting, with online casino gambling representing a large majority of patients seeking treatment at a gambling disorder facility in Sweden [16].

Sweden has a license gambling market since 2019, where online and land-based gambling operators are allowed to operate, provided they adhere to a number of responsible gambling regulations, one of which is adherence to the nationwide Spelpaus service. The legal gambling age in Sweden is 18 years [6]. After being a more traditional, land-based gambling market, but with a large predominance of a government-owned gambling monopoly for many years [21], the Swedish gambling market was gradually increasingly affected by overseas online operators, which represented a large proportion of visible gambling advertising despite their unregulated status [22]. This led to the decision to liberalize the gambling market into a licensed, but controlled, market from 2019. Given the new license-based gambling market since 2019, and thereby the relative recency of a unique self-exclusion service, the present setting is of interest to assess the qualities and challenges of such a new harm reduction instrument.

Potential stakeholders and users of the study's findings may include policy makers, public health officials, mental health professionals, gambling addiction counselors, researchers in the field of addiction studies, advocacy groups, and individuals and families affected by problem gambling. The insights garnered from this research could inform the development of targeted interventions, public health campaigns, regulatory measures, and support services aimed at addressing and mitigating the adverse effects of gambling behavior within the Swedish population.

Recruitment Process

Criterion sampling was used in the recruitment process. Participants were selected based on 2 critera: (1) being 18 years old or older during the time of the interview and (2) having previous or ongoing experience of using Spelpaus (any experience was enough, and it was not a requirement that the self-exclusion be in close temporal association with the interview). Recruitment was conducted online through social media advertisements managed by Trialy, a company specializing in research recruitment. Potential participants received information about the study and registered their interest via the Trialy website, completing a short online survey regarding their age and experience with Spelpaus. Eligible individuals were contacted by the first author (JT) and provided with additional study information. They were then sent written information about the study and about participants' rights and asked to sign and return (by mail) an informed consent form. When written consent was obtained, an interview was scheduled. Individuals were recruited during spring-summer 2023.

Sample

In total, 13 men and 2 women, aged between 31 and 62 years, participated in the study, representing various regions of Sweden. All participants reported either a gambling problem or identified themselves as being at risk, with prior or current experience using the Spelpaus self-exclusion service. This gender distribution aligned with existing research findings on gambling demographics in Sweden, which indicate a higher prevalence of gambling-related issues among men [16]. The gambling types involved in each person's gambling pattern were reported. Although the study did not systematically record whether gambling occurred (and had occurred) in online-based or land-based settings, the distribution of gambling types reported by the participants was comparable with the gambling types typically seen in Swedish problem gambling treatment settings, where online casino and sports betting is predominant [16].

Interviews

All interviews were digital (both video and audio) and were carried out by the first author (JT) of the paper, who has experience in qualitative interviewing and psychiatric research, especially in psychiatric research, ensuring a sensitive and nuanced data collection approach. The interviews were audiorecorded but not videorecorded and lasted between 23 and 67 minutes. During the interview process, the authors continuously reviewed and assessed the interviews to discern recurring patterns. Following the completion of 12 interviews, the emergence of new categories began to decline. In the final 2 interviews, only categories closely aligned with those identified previously surfaced. Consequently, the authors made the decision to conclude recruitment for the study. All interviews were then transcribed verbatim by the first and second authors (JT and SH).

As compensation for their time and efforts in the study, all participants received 2 cinema tickets.

Researcher Characteristics and Reflexivity

Our research team included professionals in psychiatry and social work, with substantial experience in qualitative research and interviewing. Throughout the study, we reflected on our backgrounds and preconceptions, particularly regarding gambling disorders and harm reduction strategies, and their potential influence on data interpretation. This reflexivity was integrated into our coding and analysis discussions to include

multiple perspectives and thereby to expand our own perspectives.

To ensure trustworthiness, triangulation was used throughout the analysis process. Multiple researchers independently coded the data, followed by discussions to reach consensus, minimizing the influence of individual biases and ensuring comprehensive data representation.

Data Analysis

This study used qualitative content analysis as the primary method for data analysis, chosen for its flexibility in systematically describing textual data, aligning well with the exploratory nature of this research. The analysis focused on manifest content, which involves identifying clear, descriptive categories that summarize data without deeper interpretive meanings. As noted by Graneheim and Lundman [23], categories represent the descriptive level of content and express the manifest content of the text, consistent with this study's goals. We deliberately chose not to identify themes, aiming instead to capture the descriptive aspects of participants' experiences, a recommended approach when focusing on manifest content rather than latent content [24,25]. Although themes can provide interpretative layers, we chose not to use them in this analysis, maintaining a descriptive focus to capture clear, observable patterns in participants' experiences with Spelpaus. This approach reflects the strength of qualitative content analysis in emphasizing the descriptive aspect without requiring thematic depth when the aim is to focus on manifest content [23].

The qualitative content analysis followed Graneheim and Lundman's [23] approach, focusing on identifying categories that represented the manifest content of participants' responses. Key analysis steps included repeated readings of the transcripts by 3 authors (JT, SH, and HH), identifying meaning units, condensing those units, and labeling them with codes, which were then grouped into categories based on similarities. The transcribed interviews were read through repeatedly and independently, and then the material was discussed by all authors. The content of the interviews related to the aim of the study was divided into meaning units, which were then condensed and labeled with codes. Similarities and differences in the codes were identified before they were divided into different categories and subcategories (Table 1). All steps of the analysis process were carried out separately and then discussed within the group to reach consensus. Various meetings within the group were held during the analysis process to discuss the different findings before a consensus about the results was reached.

Table 1. Examples of the data analysis process.

Meaning unit	Condensed meaning units	Code	Category	Subcategory
"If you should have a gambling stop like this, it should really apply to everything, because the brain will always find a way otherwise. This is the good thing—that it is a stop. You can't make a phone call, you can't just change your mind, because now it is a stop. I think that is very important." [Male participant, 31 years old]	The gambling stop should really apply to everything. Because the brain will always find a way otherwise. The good thing is that it is a stop, you cannot just change your mind. I think that is important.	Satisfaction with Spelpaus	Positive experiences	Prevention of future gambling problems
"Well, the downside is that it gives a little bit of a false sense of security. At least when it comes to certain people...as if you gamble on other gambling sites...it doesn't help anything." [Female participant, 35 years old]	The downside is that it gives a little bit of a false sense of security. When it comes to certain people who gamble on other sites, it does not help anything.	Suggestion for improvement: to also stop overseas gambling sites	Suggestions for improvement	Addressing loopholes

Ethical Considerations

The study was conducted according to the principles of the Declaration of Helsinki and was approved by the Swedish Ethical Review Authority (approval number 2022/06933-01, amendment 2023/01684-02). All participants received oral and written information about the study and about participants' rights, prior to signing a consent form. Written informed consent was obtained from all participants.

Results

Participant Details

A detailed description of the 15 participants is presented in Table 2.

Table 2. Background information about the participants (N=15).

Characteristics	Participants, n (%)
Social relationships	
In a relationship	8 (53)
Have children	9 (60)
Gambling form	
Casino	3 (20)
Sports betting	1 (7)
Horse race betting	0
Lottery	1 (7)
Casino + sports betting	2 (13)
Casino + poker	3 (20)
Casino + horse race betting	1 (7)
>2 different gambling forms	4 (27)
Enrolled in Spelpaus at the time of the interview	
Yes	10 (67)
No	3 (20)
N/A[a]	2 (13)
In active gambling at the time of the interview	
Yes	3 (20)
No	11 (73)
N/A	1 (7)
Number of times using Spelpaus	
1	4 (27)
≥2	11 (73)
Chosen time interval of self-exclusion	
12 months	7 (47)
1, 3, or 6 months	3 (20)
Both of the above	5 (33)
Overseas gambling during Spelpaus	
Yes	6 (40)
No	9 (60)

[a]Not applicable.

Categories and Subcategories

As shown in Table 3, we identified 3 categories: motivations and reasons for self-exclusion (including triggers, such as financial strain and concern for close relationships), perceived positive experiences (eg, reduced gambling urges and improved quality of life), and challenges and suggestions for improvement (notably, breaches through unlicensed gambling sites and lack of integrated support).

Table 3. Categories and subcategories.

Category	Subcategories
Reasons for the decision to self-exclude	• Precipitating factors • Self-awareness of gambling habits • Concern for close ones
Positive experiences	• Prevention of future gambling problems • Improved quality of life
Suggestions for improvement	• Integrated support systems • Addressing loopholes • Practical enhancements

Reasons for the Decision to Self-Exclude

The motivations underlying the choice to use the self-exclusion service were multifaceted, often encompassing a combination of factors.

Precipitating Factors

Numerous participants reported specific events that precipitated heightened gambling activity, ultimately leading to the decision to self-exclude. Common triggers included winning a large amount of money and quickly gambling it away, traumatic family incidents, interpersonal conflicts, financial struggles, and the onset of the COVID-19 pandemic. For instance, 1 (7%) participant articulated:

> *I was facing unemployment without insurance, which exacerbated my gambling tendencies. Although I secured employment thereafter, the meager income did not deter my gambling. The resurgence of COVID-19 in September restarted my gambling habits, leading to a downward spiral.* [Male participant, 44 years old]

Additionally, some participants opted for self-exclusion to prevent the risk of others accessing their gambling accounts as part of gambling-related criminal behavior. An example of this was the following interviewee who had a gambling problem but who also had a family member with a severe gambling problem:

> *...At the same time, this has helped us a lot every time she has wished to gamble on our accounts.* [Female participant, 35 years old]

Another reason to self-exclude was because the individual wished to facilitate debt settlement, as self-exclusion was a prerequisite for debt relief applications.

> *I applied for a debt settlement, but since I did not take it [self-exclusion] for a year, I did not get it…I would like to make another attempt for a debt settlement application, but then I probably have to take a 12-month period [of self-exclusion] as well.*

Self-Awareness of Gambling Habits

A prevalent reason for self-exclusion was an enhanced awareness of one's gambling behavior. Participants acknowledged excessive gambling tendencies, albeit with varying perspectives on whether they had developed a full-fledged gambling problem or were teetering on the brink of one. Fueled by these insights, individuals opted for Spelpaus to forestall further progression into problem gambling or to address existing issues. As 1 (7%) participant reflected:

> *I have had a realization, particularly in recent months...I recognize that I need to address this behavior before it escalates.* [Female participant, 56 years old]

This introspective process often unfolded gradually, sometimes facilitated by professional guidance or peer support.

Concern for Close Ones

Many participants described the adverse impact of their gambling habits on loved ones, with 10 (67%) individuals mentioning that their loved ones had been negatively impacted to a great extent. Financial strain, deceit, and compromised relationships were common themes. Some participants lamented fractured relationships resulting from their gambling, while others acknowledged the toll gambling took on family bonds and quality time spent together. One participant expressed:

> *It is not just about the money lost—it is the betrayal, the lies...and the time squandered gambling instead of being with my family that truly stings.* [Male participant, 39 years old]

Although self-exclusion was often an independent decision, familial influence or shared deliberation with loved ones also played a role. The following 2 (13%) citations are examples of this:

> *The relationship with my parents is starting to get a little bit better; for many years, it was pretty bad. They have helped me out with money a lot of times. Every time I call them, they think I am going to ask them for money again, so it [gambling] has hurt our relationship a lot.* [Male participant, 44 years old]

> *If you fail in a game or something, you get kind of aggressive, depressed, and angry. You also get very affected emotionally. If you bet on 7 games and 1 of them fails, you get upset, angry, or sad. It has affected my close ones the most.* [Male participant, 40 years old]

Positive Experiences

Participants identified several benefits associated with using the Spelpaus self-exclusion service.

Prevention of Future Gambling Problems

Self-exclusion was perceived as a preemptive measure to safeguard against relapse into gambling behaviors. Even individuals who had not experienced gambling urges for an extended period recognized the potential for future vulnerability. By enacting self-exclusion, participants closed the door to impulsive gambling, thus mitigating the risk of relapse. Several participants credited the Spelpaus service for preventing immediate urges to gamble. A couple of the participants who began gambling again described how they after the end of the self-exclusion period managed to gamble in a more controlled way than before.

> *After the self-exclusion, it was much more under control. I gambled a lot less, bet smaller amounts…if I felt I need to limit it now or I have to stop now. When you feel it is starting to be too much or take too much time, you feel a little self-control. Then I take a free weekend or several days free from betting, and after that, I bet a little, and then it feels like everything is under control.* [Male participant, 40 years old]

Another participant described the partial effect on gambling, primarily on impulse-driven uncontrolled gambling:

> *A part of me does not really want to say goodbye and never gamble again, but I want it to be at a reasonable level. It is at ATG [Swedish horse-betting company]. I am staying there. I think I gamble for 96 (Swedish) kronor a week now, but it was still that break (self-exclusion) that forced me to deal with it somehow.* [Male participant, 55 years old]

Improved Quality of Life

Many participants reported enhanced well-being during the self-exclusion period. Freed from the grip of gambling, individuals experienced a newfound sense of calm, happiness, and control over their lives. Moreover, self-exclusion provided an opportunity for personal growth and self-improvement, allowing participants to redirect their focus toward constructive pursuits, such as physical exercise and social interactions. One individual reflected:

> *You can relax because you have given away the control to something else…once I have chosen to self-exclude, I cannot do anything about it…and I get so mentally relaxed, I do not have to think about it at all anymore.* [Male participant, 32 years old]

Financial stability and restored trust within familial relationships were also cited as positive outcomes, as 1 (7%) respondent expressed:

> *My family can tell that I feel better when I do not gamble. I am happier, more positive, and I do not call them and ask for help. Of course, they notice a big difference when I do not gamble.* [Male participant, 31 years old]

In addition, the self-exclusion period provided an opportunity for reflection within the motivational process:

> *When I had my Spelpaus, I became quite aware of the problem, so I had a lot of thoughts about the need to do something about this problem.* [Male participant, 40 years old]

Suggestions for Improvement

Participants identified several areas for enhancement within the Spelpaus self-exclusion service.

Integrated Support Systems

Acknowledging self-exclusion as 1 component of the multifaceted approach to gambling cessation, participants emphasized the need for additional support mechanisms. One participant expressed:

> *My fear is that when the self-exclusion disappears, I have not come so far that I will not get stuck in gambling again…I think it is important that you try to solve your problems when you are self-excluded.* [Male participant, 43 years old]

Peer support groups, counseling services, and psychotherapy were deemed vital complements to self-exclusion. However, participants expressed reluctance to seek out support independently, underscoring the importance of proactive outreach and integrated support systems from within Spelpaus. Some of the participants who considered themselves to be more at risk of developing a gambling problem expressed that Spelpaus alone was enough for them.

> *If you take a year, and you have the opportunity to end it, then there needs to be some kind of dialogue before, and some kind of amount limit.* [Male participant, 55 years old]
>
> *I think it should be combined with a guaranteed CBT [cognitive behavioural therapy] or diary where you write what you feel, what you can do and what you think you should be able to do. Then of course, in the best of worlds, group therapy is perhaps even better than being digital.* [Male participant, 55, years old]

Addressing Loopholes

Participants raised concerns regarding the efficacy of Spelpaus in preventing access to overseas gambling sites. The absence of coverage for nondomestic operators posed a significant loophole, enabling some individuals to circumvent self-exclusion measures. Suggestions for improvement included enhanced international cooperation and stricter regulations on gambling advertisements.

Practical Enhancements

Participants proposed various technical and logistical improvements to Spelpaus, including more flexible and expanded self-exclusion options, longer time intervals, and a more deliberative process for ending self-exclusion periods. Mainly, participants asked for time intervals longer than 12 months. Several participants saw it as problematic to be able to gamble immediately after the end of a self-exclusion period. Additionally, participants advocated for heightened demands on gambling operators to deter relapse and streamline the self-exclusion process.

> *The self-exclusion service should be able to be developed so you cannot access overseas sites…when*

> *you select it, it should be covering all sites within the EU [European Union] or something like that…that you have a self-exclusion service for all of Europe when you choose it.* [Male participant, 62 years old]
>
> *The best thing would have been if there was a self-exclusion site where you are self-excluded from all gambling…that you have some kind of cooperation with other countries and everything and that it will be a stop everywhere.* [Female participant, 35 years old]

Several respondents stated that the possibility of returning to gambling after the self-exclusion period should still be hindered or slowed down in order to prevent impulse-related relapses even after the predetermined self-exclusion period is over. Some examples of this was an "embargo" period during which gambling still could not be reopened. Some participants found themselves waiting to gamble immediately after the self-exclusion period ended, so they suggested there be some kind of deposit limit for all clients coming back from self-exclusion, a mandatory telephone call or other message to individuals for whom self-exckusion is ending, prohibition of direct advertising to individuals immediately after a self-exclusion period, or mechanisms making the risk of gambling during alcohol influence smaller (eg, through prohibition of gambling during the night).

One respondent described the need for an embargo period:

> *After 12 months, you should not be able to cancel it on the same day; it should be 1, 2, 3 months notice…otherwise, when 12 months have passed, you can just gamble without any restrictions.* [Male participant, 32 years old]

Other respondents described the need for other possible measures to apply in the policy of self-exclusion:

> *After the self-exclusion ended, the advertising started coming again. It becomes like a reminder. So one could simply get going and gamble again!* [Male participant, 31 years old]
>
> *There should be some time to think it over when a client has asked to come back to gambling again. Some time should elapse, and then the person should be asked again. After some time, you should be asked about whether you really want to gamble again. There are people who get a kick, but after some time, that feeling migh pass again.* [Male participant, 44 years old]

Discussion

Principal Findings

This study aimed to deepen the understanding of the user experience of a nationwide, multioperator self-exclusion service, beyond what is known from recent quantitative research studies with web panel members and clinical data from patients in treatment for a gambling disorder. The potential challenges of a multioperator self-exclusion service have been demonstrated in surveys, showing that primarily among individuals with highly intense gambling practices, self-exclusion is popular but often breached by its users [13,14]. However, advantages of being self-excluded, and a deeper understanding of how this is related to the challenges of the method, largely remain to be understood.

Here, this qualitative study resulted in the description of 3 overarching categories that provide a deeper understanding of the problem: the main reasons for the decision to self-exclude, the advantages perceived by the users of this method, and the users' suggestions for improvements.

The decision to self-exclude from gambling platforms is often prompted by a sudden realization of escalating gambling behavior and a subsequent desire to regain control. Although the development of a disordered gambling pattern or financial difficulties were reported to have contributed to the decision to self-exclude, it was also reported, less expectedly, that self-exclusion may also be triggered by a wish to prevent another person from gambling on one's account. This may be a less expected reason for self-exclusion and will merit further investigation, especially in relation to the role of gambling in criminal behavior. Gambling or conducting financial transactions using somebody else's identity may represent one of the features of a severe gambling problem and also constitutes a criminal act [26].

Importantly, the role of the gambler's concerned significant others emerged as one of the reasons for self-exclusion. This highlights the importance of further addressing the interplay between individuals with gambling problems and their close ones and may also facilitate treatment processes if families of the affected individual are actively involved. This finding is consistent with previous research indicating that concerned significant others can favor treatment seeking and a favorable treatment outcome in their near ones with gambling problems [27,28]. In addition, the active participation of partners in the treatment of patients with a gambling disorder can be beneficial, both to the outcome of the affected patients and to the well-being of patients and their partners [29]. Thus, altogether, concerned significant others appear to play an important role through different phases of help seeking in individuals with problem gambling, even in the earlier harm-reducing interventions that self-exclusion is meant to represent.

Furthermore, the data showed that one reason for self-exclusion may be the intention to demonstrate one's willingness to quit gambling when formally applying for a public service where this is either required or believed to have an effect favoring the application process. This type of instrumental decision to self-exclude in order to obtain specific services or to fulfill expectations of external decision makers rarely has been reported in this context before. Whether this happens as part of an individual motivational process or entirely based on the requirements from others remains to be studied.

When external factors are cited as a reason to self-exclude, whether it be the concern of family members or the requirements from a treatment setting or authority, it can be argued that the choice to self-exclude does not represent the same decisive step in a motivational process as it does when an individual self-excludes based on their own decision to stop gambling in order to prevent or alleviate symptoms of addiction. Thus, the

possibly different levels of personal commitment associated with self-excluding will be relevant to include in the theoretical framework of self-exclusion in the future.

Overall, many favorable effects from self-exclusion were cited, despite previous experience where research has highlighted some major limitations of this service. Previous studies have been conducted using quite different methodologies, relying on larger quantitative data sets from online surveys [13,14] and clinical documentation [16], and where the findings were primarily the high rates of breaching one's self-exclusion within this service. This paper, importantly, adds to the literature regarding the present type of multioperator, nationwide self-exclusion service; in previous publications from this setting, it has been highlighted that the rate of breaching of one's self-exclusion is high in people with a gambling disorder [16] or with intensive online gambling patterns [14]. Here, several respondents described a large favorable impact on life after having self-excluded. Therefore, the description of more favorable experiences from this study provides a valuable experience, which contributes to the overall picture of this model, and strongly deepens the understanding of how this service fairs in affected individuals, over and above previous analyses conducted with quantitative methodology.

Although self-exclusion offered a respite from gambling-related stressors, several suggestions for improvements of the system emerged from the interviews. Participants emphasized the importance of holistic support systems and regulatory measures to augment its efficacy as a harm reduction tool. Suggestions for improvement partly followed what could be expected from previous web survey studies on gamblers; as many individuals who self-exclude also struggle with the risk of relapsing on overseas and nonlicensed gambling sites, users desired further legislative efforts making it possible to also self-exclude from other services, including those that are registered outside the present jurisdiction. Gambling in unlicensed, overseas online casinos has been reported to be the most common means of breaching one's self-exclusion period, as reported in the online surveys previously conducted here in this setting [13,14]. Thus, one major implication of this is the political process of how to exclude unlicensed gambling operators from the gambling market or to prevent financial transactions to gambling companies operating abroad. The legal framework and technical boundaries of this go beyond the scope of this study, but these findings highlight the need for gambling regulations to stretch beyond the controlling of the land-based gambling opportunity within each geographical jurisdiction.

One important and relatively novel suggestion was the linking between a mere harm-reducing self-exclusion service and actual therapeutic efforts against the addictive disorder. Importantly, this would move the purpose of the self-exclusion service from being an anonymous electronic support tool to actually providing a conduit to treatment. Treatment seeking in gambling disorder is known to be low, with several barriers reported [18]. Importantly, it has been reported that despite perceived barriers against treatment seeking, individuals with gambling problems may use more informal ways to seek help and not necessarily formal enrolment into psychotherapeutic treatment. For example, this may involve access to online advice or other online tools to facilitate a behavioral change [10]. In that aspect, any degree of referral from self-exclusion to different degrees of online support, advice, or treatment may be a way forward to improve the outcome of a person with gambling problems, over and above the effect of self-exclusion per se.

Such a system of transferring self-excluding clients to addiction treatment has not, to the best of our knowledge, been described in previous literature, and this may suggest a new area of research. For example, in such a system, it would have to be considered whether there is a potential barrier against enrolling in a self-exclusion service if it may, under some conditions, trigger a personal contact from a gambling operator, from a public authority, or from a treatment provider. However, it has been demonstrated that unsolicited motivational telephone calls to people displaying hazardous gambling habits may be a way to lower wagering and increase harm-reducing measures [30,31] and that acceptability of such motivational efforts appears to be high [32]. Thus, if self-exclusion from gambling would trigger an outreach effort aiming to improve the individual's gambling, it could be argued that this would not require a formal therapist-guided treatment but possibly may involve a brief, personalized intervention but with the capacity to refer individuals to further treatment when needed. Such individualized, personalized normative feedback has been suggested to be effective in other treatment or harm reduction interventions in problem gambling [33,34] and can initiate or enhance an internal motivational process in affected individuals, over and above the effects from the sole self-exclusion.

Likewise, another expansion of the self-exclusion tool, as suggested by this study, would be to slow down the procedure of opening up gambling opportunities after the self-exclusion period is terminated. This would also be a novel strategy that would require further study, but potentially, given the loss of the control component in gambling disorder, a more gradual onset of gambling after the self-exclusion period may provide further possibility to consider one's choice to gamble and may reduce the harm potentially associated with a sudden onset of high-risk gambling. Likely, any tool that slows down access to online gambling may potentially have a harm-reducing and relapse-preventing effect [35].

Overall, including the considerations made by interviewees in this study and their suggestions for improvement, a self-exclusion service generally appears to be promising and possible to increase even further. However, the major obstacle to this method, the risk of breaching one's self-exclusion through the use of gambling sites outside of the system, may remain. Here, one also has to consider the fact that the choice to self-exclude may cause the gambling patterns upon a potential relapse to happen on an unlicensed market and thereby with a lower degree of control than inside a licensed market. Previous experience, however, has pointed out that in individuals with extremely intense gambling patterns, such gambling patterns may happen on many operators, and a limit set upon one's gambling in one setting may be followed by gambling on another site. Still, even when this happens, such as during a deposit limit imposed during the COVID-19 pandemic in Sweden, still many clients report such a limit to be favorable and to have decreased their gambling pattern, rather than the

opposite [36]. Thus, although a method such as self-exclusion cannot hinder all types of gambling, it may still have motivational effects that limit the money lost and limits some of the consequences even in a high-risk gambling behaviour.

Strengths and Limitations

This study has a number of potential limitations, but given the novelty of the research topic and the merits of an in-depth qualitative study design, it also has considerable strengths. Recruitment via social media advertisements and digital interviews can be considered as both a strength and a limitation. The recruitment method enabled recruitment without any geographical limitations, therefore enabling a wide sample of individuals with different backgrounds, gambling habits, and experiences. However, it is possible that advertisements in social media attract interviewees with slightly different views on and experiences of self-exclusion from gambling than help-seeking individuals or individuals currently attending some type of land-based or online-based gambling venue. In contrast, the capacity of a qualitative design to identify a range of different aspects of a behaviour may benefit from this type of broad recruitment.

Digital interviews may represent another limitation, as these cannot readily be compared with physical interviews. Small changes in tone and body language are easily overlooked when the interview is held online. Meanwhile, the use of digital interviews in this study substantially facilitated data collection, widened the possibility to participate from the whole country rather than only locally, and may have facilitated recruitment of affected individuals who may not readily disclose their personal suffering and show up in person in a mental health–and addicion-oriented research unit. Indeed, challenges, but also potential advantages, of online qualitative research interview have been discussed in the research literature in recent years [37], pointing to the fact that such online interviews can play an important role in qualitative research.

Another potential limitation is the time aspect; some of the participants had recent or even ongoing experiences of Spelpaus, while others had been using it further back in time. Although this may be seen as a limitation, their information about their Spelpaus-related experience still is of great value (ie, as their ongoing or previous experience).

Additionally, the absence of a diagnosis of gambling gisorder as an inclusion criterion means that assessments of participants' gambling problems relied solely on self-reported statements and estimations, which may not capture the full spectrum of gambling-related issues. This limitation underscores the possibility that experiences with Spelpaus may vary based on individuals' level of gambling problems.

However, this study addressed a novel research issue and has strengths in a number of aspects. This is, to the best of our knowledge, the first qualitative study centered around the users' experiences of the relatively novel and unique nationwide self-exclusion service. Consequently, it offers updated and in-depth insights into gamblers' experiences with Spelpaus, elucidating its advantages and challenges as a harm reduction tool.

Conclusion

This qualitative study aimed to deepen the knowledge about and understanding of a novel nationwide service for self-exclusion from gambling. This includes the experience of individuals who have a history of being self-excluded, including their reasons for self-exclusion and their experience of the system's merits and challenges. Based on the interviews, this study suggests that although breaching one's self-exclusion remains a challenge, users of this service also report major advantages from being self-excluded. Users suggest further improvements to the self-exclusion model, such as international collaboration in order to prevent overseas gambling operators offering gambling outside of the self-exclusion system and also structured links from self-exclusion to further help and support, possibly also involving formal treatment. The study provided new information about self-exclusion and about a broader ranges of reasons for self-excluding, over and above the reporting of severe gambling problems (eg, the role of self-exclusion in the relationship with concerned significant others or to prove willingness to quit gambling). Several favorable effects from self-excluding were reported, in addition to a need for clearer obstacles against going back to gambling once self-exclusion ends.

Conflicts of Interest

AH has research funding for other projects from AB Svenska Spel, which is the state-owned gambling operator of Sweden, and from Svenska Spel's independent research council, as well as from the research council of Systembolaget, the Swedish alcohol monopoly. None of these organizations had any role in, or influence on, this project. The project itself was carried out thanks to funding from the research council of Svenska Spel. The funder did not have any influence on the procedures of this research or on the interpretation and reporting of results.

References

1. Langham E, Thorne H, Browne M, Donaldson P, Rose J, Rockloff M. Understanding gambling related harm: a proposed definition, conceptual framework, and taxonomy of harms. BMC Public Health 2016 Jan 27;16(1):80 [FREE Full text] [doi: 10.1186/s12889-016-2747-0] [Medline: 26818137]
2. Potenza MN, Balodis IM, Derevensky J, Grant JE, Petry NM, Verdejo-Garcia A, et al. Gambling disorder. Nat Rev Dis Primers 2019 Jul 25;5(1):51. [doi: 10.1038/s41572-019-0099-7] [Medline: 31346179]

3. Wardle H, Reith G, Langham E, Rogers RD. Gambling and public health: we need policy action to prevent harm. BMJ 2019 May 08;365:l1807 [FREE Full text] [doi: 10.1136/bmj.l1807] [Medline: 31068335]
4. Matheson FI, Hamilton-Wright S, Kryszajtys DT, Wiese JL, Cadel L, Ziegler C, et al. The use of self-management strategies for problem gambling: a scoping review. BMC Public Health 2019 Apr 29;19(1):445 [FREE Full text] [doi: 10.1186/s12889-019-6755-8] [Medline: 31035978]
5. Kraus L, Loy JK, Bickl AM, Schwarzkopf L, Volberg RA, Rolando S, et al. Self-exclusion from gambling: a toothless tiger? Front Psychiatry 2022 Sep 23;13:992309 [FREE Full text] [doi: 10.3389/fpsyt.2022.992309] [Medline: 36213894]
6. Ministry of Finance. Gambling Act (2018:1138). Swedish Gambling Authority. URL: https://www.spelinspektionen.se/globalassets/dokument/engelsk/oversatt-spellagen/english-spellagen-sfs-2018_1138-uppdat-sfs-2024_255.pdf [accessed 2025-01-21]
7. Abbott M, Romild U, Volberg R. The prevalence, incidence, and gender and age‐specific incidence of problem gambling: results of the Swedish longitudinal gambling study (Swelogs). Addiction 2017 Nov 24;113(4):699-707. [doi: 10.1111/add.14083]
8. Gabellini E, Lucchini F, Gattoni ME. Prevalence of problem gambling: a meta-analysis of recent empirical research (2016-2022). J Gambl Stud 2022 Dec 31;39(3):1027-1057. [doi: 10.1007/s10899-022-10180-0]
9. Håkansson A, Karlsson A, Widinghoff C. Treatment seeking for gambling disorder in nationwide register data – observations around a major shift in legislation. Front Public Health 2024 Mar 19;12:1293887 [FREE Full text] [doi: 10.3389/fpubh.2024.1293887] [Medline: 38566789]
10. Rodda S, Dowling N, Lubman D. Gamblers seeking online help are active help-seekers: time to support autonomy and competence. Addict Behav 2018 Dec;87:272-275. [doi: 10.1016/j.addbeh.2018.06.001] [Medline: 29935737]
11. Exclude yourself from gambling!. Spelpaus. 2024. URL: https://www.spelpaus.se/en#:~:text=This%20is%20where%20you%20can,real%2Dmoney%20gambling%20in%20Sweden.&text=A%20self%2Dexclusion%20via%20Spelpaus,before%20the%20date%20you%20choose [accessed 2025-01-21]
12. Håkansson A, Widinghoff C, Berge J. Self-exclusion from gambling – a measure of COVID-19 impact on gambling in a highly online-based gambling market? Int J Environ Res Public Health 2021 Jul 09;18(14):7367 [FREE Full text] [doi: 10.3390/ijerph18147367] [Medline: 34299817]
13. Håkansson A, Widinghoff C. Gambling despite nationwide self-exclusion - survey in online gamblers in Sweden. Front Psychiatry 2020 Dec 2;11:599967 [FREE Full text] [doi: 10.3389/fpsyt.2020.599967] [Medline: 33343428]
14. Håkansson A, Komzia N. Self-exclusion and breaching of self-exclusion from gambling: a repeated survey study on the development of a nationwide self-exclusion service. Harm Reduct J 2023 Aug 08;20(1):107 [FREE Full text] [doi: 10.1186/s12954-023-00822-w] [Medline: 37553657]
15. Spelvanor online 2023 [Online gambling habits 2023]. Swedish Gambling Authority. URL: https://www.spelinspektionen.se/globalassets/dokument/statistik/enkatundersokning/spelvanor-online-rapport-juli-2023.pdf [accessed 2025-01-21]
16. Miles M, Rothschild J, Åkesson G, Håkansson A. Nationwide, multioperator self-exclusion and psychiatric comorbidity in patients with gambling disorder: a retrospective chart review study from a regional treatment unit. J Addict 2023;2023:5532259 [FREE Full text] [doi: 10.1155/2023/5532259] [Medline: 37808466]
17. LaPlante DA, Nelson SE, LaBrie RA, Shaffer HJ. Stability and progression of disordered gambling: lessons from longitudinal studies. Can J Psychiatry 2008 Jan 01;53(1):52-60. [doi: 10.1177/070674370805300108] [Medline: 18286872]
18. Suurvali H, Cordingley J, Hodgins DC, Cunningham J. Barriers to seeking help for gambling problems: a review of the empirical literature. J Gambl Stud 2009 Sep 24;25(3):407-424. [doi: 10.1007/s10899-009-9129-9] [Medline: 19551495]
19. Håkansson A, Tjernberg J, Hansson H. Effects and limitations of a unique, nationwide, self-exclusion service for gambling disorder and its self-perceived effects and harms in gamblers: protocol for a qualitative study. JMIR Res Protoc 2023 Nov 14;12:e47528 [FREE Full text] [doi: 10.2196/47528] [Medline: 37962917]
20. Public Health Agency of Sweden. Swelogs (The Swedish Longitudinal Gambling Study). Public Health Agency of Sweden: Sweden; 2023.
21. Binde P. Gambling in Sweden: the cultural and socio-political context. Addiction 2014 Feb;109(2):193-198. [doi: 10.1111/add.12103] [Medline: 23297702]
22. Håkansson A, Widinghoff C. Television gambling advertisements: extent and content of gambling advertisements with a focus on potential high-risk commercial messages. Addict Behav Rep 2019 Jun;9:100182 [FREE Full text] [doi: 10.1016/j.abrep.2019.100182] [Medline: 31193826]
23. Graneheim U, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
24. Jacobsen D. Förståelse, beskrivning och förklaring - Introduktion till samhällsvetenskaplig metod för hälsovård och socialt arbete. Lund: Studentlitteratur; 2012.
25. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
26. Binde P, Cisneros Örnberg J, Forsström D. Criminogenic problem gambling: a study of verdicts by Swedish courts. Int Gambl Stud 2021 Dec 08;22(3):344-364. [doi: 10.1080/14459795.2021.2002383]

27. Grant JE, Kim SW, Kuskowski M. Retrospective review of treatment retention in pathological gambling. Compr Psychiatry 2004 Mar;45(2):83-87. [doi: 10.1016/j.comppsych.2003.12.005] [Medline: 14999657]
28. Ingle PJ, Marotta J, McMillan G, Wisdom JP. Significant others and gambling treatment outcomes. J Gambl Stud 2008 Sep 5;24(3):381-392. [doi: 10.1007/s10899-008-9092-x] [Medline: 18320296]
29. Tremblay J, Dufour M, Bertrand K, Saint-Jacques M, Ferland F, Blanchette-Martin N, et al. Efficacy of a randomized controlled trial of integrative couple treatment for pathological gambling (ICT-PG): 10-month follow-up. J Consult Clin Psychol 2023 Apr;91(4):221-233. [doi: 10.1037/ccp0000765] [Medline: 36521131]
30. Jonsson J, Hodgins DC, Munck I, Carlbring P. Reaching out to big losers leads to sustained reductions in gambling over 1 year: a randomized controlled trial of brief motivational contact. Addiction 2020 Aug 13;115(8):1522-1531. [doi: 10.1111/add.14982] [Medline: 31977104]
31. Hakansson A, Franklin K, Dahlström M, Lyckberg A. Motivational telephone intervention to risk gamblers by a state-owned gambling operator in Sweden. Front Psychiatry 2024 Jan 30;15:1343733 [FREE Full text] [doi: 10.3389/fpsyt.2024.1343733] [Medline: 38352656]
32. Håkansson A, Franklin K, Dahlström M, Lyckberg A. Responsible gambling through a motivational telephone intervention to high-risk gamblers - an evaluation of user satisfaction and subjective intervention effects. Front Psychol 2022 Nov 30;13:917216 [FREE Full text] [doi: 10.3389/fpsyg.2022.917216] [Medline: 36532977]
33. Neighbors C, Rodriguez LM, Rinker DV, Gonzales RG, Agana M, Tackett JL, et al. Efficacy of personalized normative feedback as a brief intervention for college student gambling: a randomized controlled trial. J Consult Clin Psychol 2015 Jun;83(3):500-511 [FREE Full text] [doi: 10.1037/a0039125] [Medline: 26009785]
34. Marchica L, Derevensky JL. Examining personalized feedback interventions for gambling disorders: a systematic review. J Behav Addict 2016 Mar;5(1):1-10 [FREE Full text] [doi: 10.1556/2006.5.2016.006] [Medline: 28092190]
35. Newall PWS. Reduce the speed and ease of online gambling in order to prevent harm. Addiction 2023 Feb 24;118(2):204-205. [doi: 10.1111/add.16028] [Medline: 36002979]
36. Håkansson A, Sundvall A, Lyckberg A. Effects of a national preventive intervention against potential COVID-19-related gambling problems. A self-report survey in online gamblers. JMIR Form Res 2022 Mar 09;6(3):e33066 [FREE Full text] [doi: 10.2196/33066] [Medline: 34678751]
37. Keen S, Lomeli-Rodriguez M, Joffe H. From challenge to opportunity: virtual qualitative research during COVID-19 and beyond. Int J Qual Methods 2022 Jun 04;21:16094069221105075. [doi: 10.1177/16094069221105075] [Medline: 35692956]

Edited by A Kushniruk; submitted 02.09.24; peer-reviewed by M Bellringer, R Lantto; comments to author 14.10.24; revised version received 23.12.24; accepted 08.01.25; published 31.01.25.

Please cite as:
Tjernberg J, Helgesson S, Håkansson A, Hansson H
Exploring the Users' Perspective of the Nationwide Self-Exclusion Service for Gambling Disorder, "Spelpaus": Qualitative Interview Study
JMIR Hum Factors 2025;12:e66045
URL: https://humanfactors.jmir.org/2025/1/e66045
doi:10.2196/66045
PMID:39888656

Original Paper

Barriers and Facilitators to User Engagement and Moderation for Web-Based Peer Support Among Young People: Qualitative Study Using the Behavior Change Wheel Framework

Ananya Ananya[1], MSc; Janina Tuuli[1], MSc; Rachel Perowne[1], MSc; Leslie Morrison Gutman[1], PhD

Department of Clinical, Educational and Health Psychology, Faculty of Brain Sciences, University College London, London, United Kingdom

Corresponding Author:
Leslie Morrison Gutman, PhD
Department of Clinical, Educational and Health Psychology
Faculty of Brain Sciences
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 8059 0939
Email: l.gutman@ucl.ac.uk

Abstract

Background: Peer support groups or web-based chats for young people offer anonymous peer support in judgment-free spaces, where users may share their thoughts and feelings with others who may have experienced similar situations. User engagement is crucial for effective web-based peer support; however, levels of engagement vary. While moderation of peer support groups can have a positive impact on the engagement of young people, effective moderation can be challenging to implement.

Objective: This study aimed to identify barriers and facilitators to user engagement with, and moderation of, web-based peer support groups among young people aged 16 to 25 years and to provide recommendations for enhancing this service.

Methods: Drawing upon the Theoretical Domains Framework (TDF) and the Behavior Change Wheel (BCW), this study conducted qualitative interviews and gathered open-ended questionnaires from service users and moderators of The Mix, the United Kingdom's leading web-based mental health platform providing peer support groups for young people. Semistructured interviews were conducted with 2 service users and 8 moderators, and open-ended questionnaires were completed by 7 service users. Themes were coded using the Capability, Opportunity, Motivation, and Behavior (COM-B) model and the TDF. The BCW tools were then used to identify relevant behavior change techniques to improve user engagement in, and moderation of, the service.

Results: Thematic analysis revealed a total of 20 inductive themes within 10 TDF domains—9 (45%) for engagement and 11 (55%) for moderation. Of these 20 themes, 3 (15%) were facilitators of engagement, 7 (35%) were facilitators of moderation, 4 (20%) were barriers to moderation, and 6 (30%) barriers to engagement. Results suggest that skills, knowledge, beliefs about consequences, intentions, emotions, and the social and physical environment are important factors influencing service users and moderators of group chats. In particular, supporting the improvement of memory, attention, and decision-making skills of those involved; adapting the physical environment to facilitate effective interactions; and reducing negative emotions are suggested to optimize the value and effectiveness of peer support groups for young people's mental health for both the service users and moderators of these services.

Conclusions: The study demonstrates the effectiveness of the BCW approach and the use of the TDF and COM-B model to understand the influences on behavior in a systematic manner, especially for mental health and well-being interventions. The findings can be applied to design structured interventions to change behaviors related to the engagement with, and moderation of, web-based peer support groups and, in turn, improve mental health outcomes for young people.

(JMIR Hum Factors 2025;12:e64097) doi:10.2196/64097

KEYWORDS

internet; moderation; engagement; youth; teenager; adolescent; peer support; web-based group; user engagement; support group; barrier; facilitator; Theoretical Domains Framework; Behavior Change Wheel; qualitative; interview; behavior change technique; thematic analysis

Introduction

Background

Young people aged 16 to 25 years are particularly vulnerable to mental health difficulties [1]. Recent evidence shows an increase in reported mental health problems for young people aged 16 to 25 years, particularly following the COVID-19 pandemic [2,3]. For example, in young people aged 17 to 19 years in England, rates of probable mental disorders rose from 17.7% in 2020 [4] to 23.3% in 2023 [5]. Nevertheless, health care systems and support often overlook the needs of young people [6]. In the United Kingdom, 75% of young people experience delays in accessing mental health support, leading to a worsening of their condition [7].

Web-based mental health communities are a viable option to bridge the mental health service gap for young people. They offer anonymous peer support in judgment-free spaces, where users may share their thoughts and feelings with others who may have experienced similar situations [8]. Web-based peer support can be asynchronous or synchronous, providing online support in a group format, with or without moderation [9,10]. Synchronous or real-time support, such as web-based chats, provides users with in-the-moment assistance without the delays that can occur within asynchronous services [11,12]. Notably, studies have shown the efficacy of web-based peer support platforms compared to in-person talk therapies [13]. They offer clinical effectiveness [14] and scalability for public health impact, enabling outreach to a larger population [13]. Therefore, there is a strong argument for the use of web-based peer support platforms for young people as they address key barriers and hold the potential to enhance mental health treatment by providing a safe, accessible, and effective means of support.

Despite numerous benefits, studies have also highlighted challenges to web-based peer mental health support that limit its effectiveness [15], including varying engagement rates [14]. Low rates of engagement can lead to negative outcomes, such as perceived exclusion and isolation [16], whereas high rates have been found to improve young people's mental health [17]. Effective moderation of web-based, user-led mental health services has been found to improve user engagement [18]. However, there are few studies examining user engagement in, and moderation of, synchronous web-based peer support groups, especially capturing the perspectives of both service users and moderators [11,19]. Service users, with their lived experience of giving and receiving support, can offer valuable insights and suggestions for improving these platforms. Moderators, due to their proximity to users and responsibility for supporting positive interactions, safety, and engagement [14], may have practical insights to enhance platform effectiveness [20]. A systematic examination of these behaviors from different viewpoints would enable the identification of tailored intervention strategies to improve moderation of web-based peer support services and enhance user engagement. Using the Behavior Change Wheel (BCW) [21], this qualitative research study investigates the barriers and facilitators to user engagement and moderation for web-based peer support groups among young people aged between 16 and 25 years and then proposes tailored strategies for optimization.

User Engagement in Web-Based Peer Support

Digital health research has defined user engagement as the extent and subjective experience (characterized by attention, interest, and affect) of use [22]. Studies measure engagement in terms of use time, log-ins, or module completion [23,24], with approximately 60% of studies measuring attendance alone to assess engagement [25]. However, although passive involvement or observational participation in web-based peer support groups has recognized benefits [26], considering only attendance falls short as it omits active participation and interaction, which fosters the sense of community and support unique to this service. Therefore, this study considers engagement in the broader sense, incorporating any observable written contribution, such as comments expressing thoughts and feelings or support for others. This inclusive approach captures subjective experience, aids barrier and facilitator identification, and can inform interventions for enhanced meaningful engagement.

Previous research highlights key barriers and facilitators to young people's engagement with digital mental health interventions and services [27,28]. Barriers include interventions that are perceived as unappealing or unhelpful, technical issues, privacy concerns, and young people lacking time or remembering to use the intervention. Facilitators include the personalization and flexibility offered by digital interventions, along with effective design, usability, opportunities to build connections, and the potential for a rewarding user experience [27,28].

However, there is a gap in research examining influences on young people's engagement in synchronous web-based peer support groups for mental health. Barriers and facilitators are likely to differ from those in other digital mental health services, as peer-to-peer support is viewed as less stigmatizing and more relatable [29]. In addition, synchronous interactions have a different pace compared to other formats [11]. Recent research focusing on peer-to-peer chats has centered on moderators' roles without including service user voice [30].

Moderation of Web-Based Peer Support

Moderation involves managing content, safeguarding, and providing support to users when needed [18]. It can be carried out by health professionals or trained volunteers [19]. Effective moderation has been found to have a positive effect on user engagement and has also been helpful in preventing "toxic" web-based settings [18]. Negative effects of web-based platforms such as trolling and stalking may also be prevented by effective moderation [16]. In web-based chat groups and

discussion forums, moderators ensure adherence to group guidelines to enhance user safety and prevent service users from becoming overly dependent on one another [31,32].

There has been limited research on the moderation of web-based mental health interventions [19]. Synchronous group settings pose distinct challenges as they require real-time monitoring of the chat and interaction between peers and moderators. In a recent study, Deng et al [30] explored moderation of peer support for a web-based mental health community. In this study, although moderators were qualified mental health professionals, they experienced challenges, such as dealing with emotionally triggering problems or hostile users. In a similar environment, Saha et al [33] highlighted challenges, such as a lack of concern for moderators' well-being and moderators' uncertainty about when to intervene. However, the services examined in these studies were not specifically designed for young people. Thus, there is a need to understand moderation and, importantly, how to enhance it, particularly in the context of real-time group chats for young people moderated by young volunteers, in a way that is structured and rooted in theory.

To address these research gaps, this study triangulates moderator and service user perspectives using a theoretically based behavior change approach, providing a comprehensive examination of user engagement and the moderation process in synchronous web-based mental health peer chats. The BCW provides a systematic and comprehensive framework to enhance understanding of behavior and support the identification of linked, relevant behavior change techniques (BCTs) to address the identified barriers and optimize group chat engagement [21].

BCW Framework

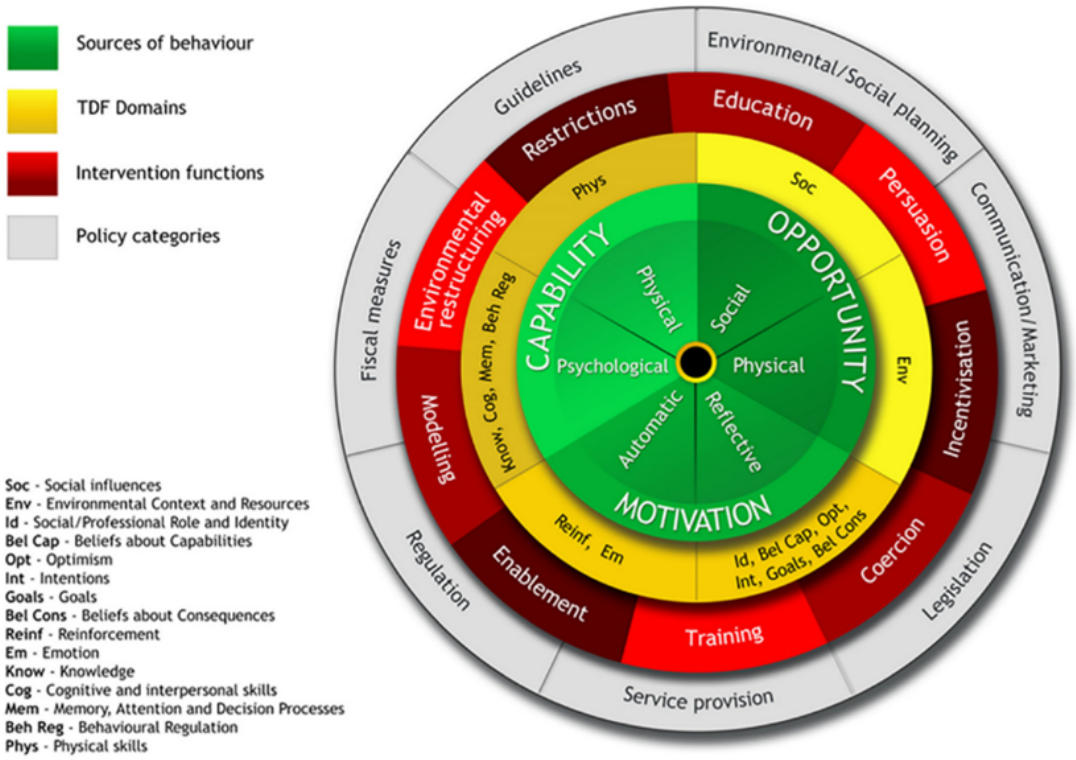
The BCW comprises interconnected tools to guide decision-making and facilitate the systematic development and evaluation of behavioral interventions [21]. At its center lies the capability, opportunity, motivation, and behavior (COM-B) model of behavior, which supports the identification of the barriers and facilitators to a behavior (Figure 1 [21]). Capability refers to an individual's physical as well as psychological capacity to perform a specific activity. Motivation is the cognitive processes that drive human behavior. Finally, opportunity refers to external factors that are outside the individual and may facilitate behavior.

The components of COM-B can be further elaborated into 14 domains using the Theoretical Domains Framework (TDF), providing a more granular understanding of the influences on behavior [34,35]. In the next layer of the BCW are intervention functions, which are the broad categories of means by which an intervention can change behavior. The outer layer consists of policy categories to support long-term, system-wide implementation.

The Behavior Change Techniques Taxonomy (BCTT) comprises 93 observable and replicable BCTs. The BCTT is a reliable method for specifying, interpreting, and implementing BCTs, which are considered "active ingredients" to facilitate behavior change either alone or in combination [36]. Expert consensus allows for mapping of the COM-B and TDF domains to intervention functions and corresponding BCTs, guiding appropriate strategies for change [37].

The BCW approach has been successfully applied in previous research to understand the barriers and facilitators to the use and delivery of digital mental health platforms for young people, including webchat counseling [28,38] and moderation of self-harm content in web-based discussion boards [39].

Figure 1. The Behavior Change Wheel and Theoretical Domains Framework (TDF) (adapted from Michie et al [21]).



This Study

Given the lack of research in this area, this qualitative study systematically explores the barriers and facilitators to the user engagement and moderation of synchronous web-based peer group chats for young people and proposes evidence-based strategies for improvement. Data were gathered from open-ended questionnaires and in-depth, semistructured interviews with young service users and moderators of The Mix, a web-based mental health support service for young people aged <25 years. Using the BCW [21], the following research questions were addressed:

1. Using the COM-B and TDF, what are the barriers and facilitators to young people's engagement in web-based peer support groups?
2. Using the COM-B and TDF, what are the barriers and facilitators to moderation of web-based peer support groups?
3. Using the BCTT, what strategies can be used to improve user engagement in, and moderation of, web-based peer support groups for young people's mental health?

Methods

Participants

This study collected data from service users and moderators of The Mix, the United Kingdom's leading digital mental health support service for young people aged <25 years. The Mix offers free mental health support, including a helpline, phone and webchat counseling, crisis support, discussion boards, and moderated live group chats. These synchronous group chats are freely available for young people aged 13 to 25 years. Table 1 summarizes the peer-to-peer mental health group chats of The Mix analyzed in this study.

The Mix circulated a poster to moderators (total population of 52) and service users to promote participation. Interested participants contacted the researcher through email, and their involvement was entirely voluntary. As shown in Table 2, a total of 2 service users were interviewed, and a further 7 service users completed an open-ended questionnaire about their engagement. A total of 8 moderators (n=5, 63% staff and n=3, 38% volunteers) participated in interviews about moderation. Of these 8 moderators, 4 (50%) staff members and 1 (13%) volunteer also completed a separate interview about user engagement. As this was a qualitative study, it aimed to gather rich and detailed data from participants, similar to previous studies in a similar context [38,39], rather than to achieve a representative sample.

All moderators were women, while most of the service users were either women or did not self-identify their gender, with the exception of 1 man and 1 nonbinary service user. Moderators were aged between 19 and 30 years, and service users were aged between 18 and 24 years (Table 2). Moderators were involved with The Mix ranging from 6 months to as long as 8 years and worked for 1 to 2 hours per week.

Semistructured interviews, based on the TDF, were conducted via the Microsoft Teams platform and transcribed for analysis. The interviews lasted between 30 to 60 minutes each and were designed to gain in-depth service user and moderator insights into experiences of group chats regarding peer interactions, structure of the chats, and feelings about contributing. Questions included, for example, "How long do you spend on The Mix's chat per week?" and "How do other people in the group chat

make it easier or more difficult for you to share your thoughts, feelings and experiences or support others with theirs?" Questions for moderators only included the following: "How do other people influence how you contribute to the moderation of posts on the group chats?"

For user engagement only, participants were offered the alternative option to complete an open-ended Microsoft Forms questionnaire to combat any difficulties young people felt in expressing themselves in an interview and may have hindered participation [28]. The questionnaire included 5 open-ended questions, exploring what made it easier or harder for users to share their experiences, what influenced their decision to participate or not, and what changes they would make to the chat if given the opportunity.

Table 1. The Mix's peer-to-peer mental health chat groups.

Group chat type	Description	Participants, n	Structure	Participation	Privacy	Moderation
Support chat	Peer support for mental health struggles	Variable	Open discussions	Active participation	No sign-up required, no password protection	Moderated by trained young volunteers and staff
Support circle	More intimate setting, where everyone takes turns to obtain support while the rest of the group listens	≤5	Rotating support and sharing	Sign up to receive support or as a "listener"	Password protected, full session attendance	Moderated by trained young volunteers and staff

Table 2. Participant information.

Participant	Data	Age (y)	Gender
SU1[a]	Interview	24	Man
SU2	Interview	18	Nonbinary
SU3	Questionnaire	18-25	Did not self-identify
SU4	Questionnaire	18-25	Did not self-identify
SU5	Questionnaire	18-25	Woman
SU6	Questionnaire	18-25	Woman
SU7	Questionnaire	18-25	Did not self-identify
SU8	Questionnaire	18-25	Did not self-identify
SU9	Questionnaire	18-25	Did not self-identify
MS1[b]	Interview (engagement and moderation)	28	Woman
MS2	Interview (engagement and moderation)	30	Woman
MS3	Interview (engagement and moderation)	24	Woman
MV4[c]	Interview (moderation)	26	Woman
MV5	Interview (engagement and moderation)	24	Woman
MV6	Interview (moderation)	19	Woman
MS7	Interview (moderation)	23	Woman
MS8	Interview (moderation)	22	Woman

[a]SU: service user.

[c]MS: moderator staff.

[d]MV: moderator volunteer.

Data Analysis

Data were thematically analyzed using the 6-phase process developed by Braun and Clarke [40]. NVivo (Lumivero) and Excel (Microsoft) were used as analytical tools. The initial stage involved gaining familiarity with the transcribed interviews. A deductive coding approach was adopted initially, which involved reviewing items of data from the transcripts and questionnaires and organizing them into the TDF domains, providing a "start list" [41,42] of themes to be categorized as barriers or facilitators. The process was repeated using an inductive analysis to generate specific subthemes within each TDF theme. These subthemes were then reviewed, defined, and included in the codebook [43]. A codebook was developed to facilitate coding and was iteratively updated. As new data were added to TDF domains, facilitators and barriers were modified, expanded, or recategorized using the constant comparison method to identify patterns and variations within the dataset [44]. Triangulation

was completed with data from both service users and moderators to explore consistencies and contradictions in influences on young people's engagement with group chats. Reflexivity was prioritized to ensure openness and challenge in interpretations [45]. To ensure coding reliability and validity, 2 researchers independently cross-coded 1 transcript; a reliability check was conducted, and agreement was reached on any discrepancies.

Following data analyses, the coded TDF barriers were mapped to corresponding intervention types, using the BCW, and then specific BCTs were selected using the links specified in Cane et al [46] as well as the more recently developed Theory and Techniques Tool [47,48]. The selection of the relevant and appropriate intervention types and BCTs was informed by an appraisal of the affordability, practicability, effectiveness and cost-effectiveness, acceptability, side effects, and equity criteria [21]. This approach has been used in a multitude of studies [49,50] to evaluate intervention strategies. The selection of policy options was out of scope because the study was focused on a single organization rather than system level. Finally, the BCTs were operationalized, and intervention strategies were proposed based on a review of previous literature.

Ethical Considerations

Low-risk ethics approval was obtained from the University College London Ethics Committee (Z6364106/2023/03/149 social research, 25069/001). Participant information and a consent form were provided to interested parties to sign digitally and confirm consent. Participants were advised that participation was completely voluntary and they could withdraw at any time, up to 4 weeks after their data were collected. As reimbursement for the interviews, participants were given a £10 (US $12.45) voucher. Questionnaire respondents could opt into a lottery prize draw to win 1 of 2 £10 (US $12.45) vouchers by providing their email address. To protect participant privacy and confidentiality, data were fully anonymized.

Results

Overview

Table 3 sets out the themes identified as both barriers and facilitators to user engagement and moderation. These are set out according to both COM-B and TDF frameworks.

Table 3. The COM-B[a] and TDF[b] themes and subthemes for user engagement and moderation.

COM-B and TDF themes	User engagement subthemes	Moderation subthemes
Psychological capability		
Cognitive and interpersonal skills	• Difficulty expressing feelings and needs (barrier) • Skills to listen and validate (facilitator)	• Developing skills (facilitator)
Knowledge	—[c]	• Understanding the guidelines (facilitator)
Memory, attention, and decision-making	• Memory recall of guidelines (barrier)	• Responding quickly in complex situations (barrier)
Behavioral regulation	—	• Holding back on advice (barrier)
Reflective motivation		
Beliefs about consequences	• Judgment and confidentiality concerns (barrier)	• Impact of moderation (facilitator)
Social and professional role and identity	—	• Congruence between social and professional identity (facilitator)
Intentions	• Wanting to support others (facilitator)	• Intending to shape young people's perceptions (facilitator)
Automatic motivation		
Emotions	• Fear and anxiety (barrier)	• Distressing subject matter (barrier)
Physical opportunity		
Environment, context, and resources	• Overload and lack of structure (barrier)	• Organizational resources (facilitator) • Lack of visual cues (barrier)
Social opportunity		
Social influences	• Similarity and familiarity (facilitator) • Integration of new users and gaps in support (barrier)	• Support offered (facilitator)

[a]COM-B: Capability, Opportunity, Motivation, and Behavior.

[b]TDF: Theoretical Domains Framework.

[c]Not applicable.

Young People's Engagement

As shown in Table 3, a total of 9 barriers and facilitators were identified as subthemes across 7 domains of the TDF. Themes were mainly consistent across support chat and support circle responses, but a small number of differences arose between the 2, and these are indicated, where relevant, in the following subsections.

Cognitive and Interpersonal Skills (Psychological Capability)

Difficulty Expressing Feelings and Needs

Several service users and moderators described the difficulty users face in expressing their emotions and thoughts, which can prevent them from opening up in the chat. Subthemes are described below, together with illustrative quotes. Service user quotes are indicated by "SUX," volunteer moderator quotes are indicated by "MVX," and staff moderator quotes are indicated by "MSX," where "X" is the participant number, to identify each participant as per Table 2. One user commented the following:

> *[What makes it more difficult to open up is] not being able to express myself about feelings.* [SU8]

Moderators also noted that vague statements about one's emotional state without specific requests for support may be a barrier to others' engagement, as they provide insufficient guidance for other users to offer relevant support. For example, one user stated the following:

> *I think it depends on how someone phrases the question...if they're sort of saying, "And I'm feeling really low today" or "And I can't do this anymore," like some of those trickier comments.* [MV5]

Skills to Listen and Validate

Both moderators and users identified that having skills to effectively validate and listen to others facilitated peer support and engagement. One moderator noted the following:

> *The people that tend to offer the most peer support are those that...know how to validate someone's experiences.* [MS1]

Users described how they did this, with one user stating the following:

> *Everyone listened and waited their turn...I listened to others, and they encourage people waiting their turn to help others.* [SU6]

Personal experiences that related to others' struggles made it easier to connect and offer support, but 1 user stated the following:

> *There are some situations where I can't exactly relate...so I stay quiet.* [SU5]

Memory, Attention, and Decision Processes (Psychological Capability): Memory Recall of Guidelines

Difficulties in absorbing, retaining, and applying guidelines posed a significant barrier to a couple of users. Users indicated that they struggled to remember guidelines due to content volume, distractions, and cognitive limitations, impacting their ability to engage. Even long-term users found it challenging, with 1 individual expressing the following:

> *I can't actually remember what [the guidelines are]. They're very basic from what I do remember.* [SU1]

Mental health crisis situations further exacerbated guideline oversight, leading some users to unintentionally deviate from the chat's intended mode of engagement in the moment. Moderators noted that this issue was worsened by some users, particularly regular ones, who found the guidelines unengaging and tended to ignore them.

Beliefs About Consequences (Reflective Motivation): Judgment and Confidentiality Concerns

This was a barrier revolving around young people's apprehensions about being judged, misunderstood, or facing negative reactions when sharing personal thoughts and experiences, particularly in the presence of new members. One participant stated the following:

> *Some people feel like a new person may judge more or less or may be offended more or be offended less. So [withdrawing] very much sort of safeguarding themselves, protecting themselves, and protecting others.* [SU1]

This could also lead users "to not give too much information" (SU7). Similarly, a moderator highlighted concerns about users fearing judgment, with 1 moderator stating the following:

> *...judging that person for talking about that cause I think that's probably what they worry about quite a lot.* [MV5]

Both moderators and users mentioned concerns about confidentiality, giving examples of where their privacy was violated by other members, leading to negative consequences. As a result, they became hesitant to share. One moderator stated the following:

> *Some members also worry confidentiality will be broken so they don't open up in fear of the police being called.* [MS2]

Intentions (Reflective Motivation): Wanting to Support Others

Overall, the intention to create a supportive environment was a facilitator, which motivated users to engage to support others. One user expressed the following:

> *Most of the time, I try my best to offer support and be there for the person too. I just want people to be able to open up to me, and when they do, I try to be a good friend.* [SU2]

However, moderators were concerned that some users may prioritize self-support, limiting their involvement in supporting others and welcoming newcomers. One moderator stated the following:

> *Often times, the young people are so busy sharing their own experiences that they don't make enough space to support others who may be going through a similar thing.* [MS1]

Emotions (Automatic Motivation): Fear and Anxiety

A frequently mentioned barrier by both users and moderators related to the user's fear of being perceived as a burden by sharing their thoughts, feelings, and experiences. One user expressed the following:

> *As for not opening up so much that's more when I feel like I'm a burden and not wanted, but that's me not The Mix.* [SU4]

Users identified that "Worrying what people think" (SU7) in terms of oversharing or repeating themselves caused them to hold back or stop sharing altogether. Another user shared similar sentiments, noting that opening up more "doesn't really make me feel better, it just makes me anxious and guilty" (SU2). These feelings of self-doubt and anxiety hindered users from fully opening up and receiving the support they needed.

Environmental Context and Resources (Physical Opportunity): Overload and Lack of Structure

A barrier identified by almost all participants was the lack of structure and organization within the support chat, especially during busy times. This could result in confusion, overwhelm, and a sense of disorganization. Users sometimes struggled to fully engage and provide or receive support effectively because, as 1 user said, "When you have a lot of people, you do get to a point where you do have like five, ten different conversations going on, all different levels of importance so it does get a bit confusing" (SU1) and "It can be hard to type that fast" (SU6). One moderator added the following:

> *It can be hard for someone to post, and their message may get missed.* [MV5]

Another user highlighted the following:

> *Support Chat needs to be more organized...The way it currently is...stresses me out, I find myself leaving midway most of the time.* [SU9]

Social Influences (Social Opportunity)

Similarity and Familiarity

The presence of individuals who had experienced similar situations or who were known to each other was highlighted by moderators and users as helping foster a sense of belonging and encouraging engagement. One user noted the following:

> *It helps that there are people who know what you've been through, so you don't feel alone. It makes you feel more comfortable to open up.* [SU1]

It was seen to be "Easier to have people with similar problems and get ideas surrounding mental health teams etc." (SU3). Several long-term users also highlighted that regular participation by both users and moderators builds trust and familiarity, facilitating engagement. One user stated the following:

> *Because we've been with [the moderators] for some time, they understand our situations a little bit more so it makes it easier for us to open up to them and...we can trust them.* [SU5]

Moderators shared similar insights, noting that regular users "feel quite attached to the moderators, so they'll leave support circle and join support chat for a bit and then come back" (MV5).

Integration of New Users and Gaps in Support

A frequently mentioned barrier by several members and most moderators related to how some regular users tended to chat among themselves or with the moderators, making it difficult for new users to integrate into the group. This could lead to new users feeling like outsiders and being less likely to engage. One user stated the following:

> *When I first started coming to circle, it was very awkward for me because the other people in there had been doing it a lot longer and I felt like a bit of an outsider.* [SU5]

New users described not receiving responses to messages, which seemed to have a notable impact on them. One of them stated the following:

> *What prompts me to withdraw is when some of my messages get ignored. I know it's not on purpose but it kind of makes me feel unwanted.* [SU4]

This was a theme that was also recognized by multiple moderators as a deficiency in social support.

Moderation

As shown in Table 3, a total of 11 barriers and facilitators were identified as subthemes across 10 domains of the TDF. The following paragraphs describe the themes in more detail, along with selected quotes.

Cognitive and Interpersonal Skills (Psychological Capability): Developing Skills

Most moderators discussed how they developed their skills over time through a continuous learning process, which helped them become more effective and refine abilities such as active listening and conflict management:

> *But it's like a continuous learning process, so I wouldn't say they give you kind of training at the beginning of you know, how to be an active listener and how to give that specific type of support.* [MS1]

Observing and practicing moderating also helped develop the necessary skills for moderation. One moderator expressed the following:

> *I think a lot of it is just sort of the exposure and just sort of the repeated attendance of chats and you sort of build, you work out sort of what things you've said have gone down well.* [MS7]

Knowledge (Psychological Capability): Understanding the Guidelines

Knowing and understanding the guidelines and procedures laid out by The Mix was a recurrent enabling theme of moderation, mentioned by all moderators, facilitating the identification of what is and is not deemed inappropriate. One moderator expressed the following:

> *And then the handbook also has things around the tech side of things, so using the platform how to use the moderator functions as well, so we're able to mute people. We're able to freeze them and then remove them from the room as well.* [MS3]

The guidelines were also particularly helpful for directing individuals at risk through appropriate safeguarding procedures. One moderator stated the following:

> *"It's got things from and sort of managing young people that come in who are in crisis. So many kinds of questions to ask safely within the room and the kind of signpost we can give.* [MS2]

Memory, Attention, and Decision Processes (Psychological Capability): Responding Quickly in Complex Situations

The complexity of group chats acted as a barrier for most moderators, presenting challenges by their fast-paced nature and situations that moderators may not have come across before. Attentively responding to the chats becomes challenging. One moderator expressed the following:

> *But there's only so fast you can type, and only so many conversations you can kind of have going on at once that you can't talk to everybody.* [MS8]

In addition, dealing with novel or gray situations caused confusion about what is acceptable in the chat and what is not, making decision-making difficult. One moderator stated the following:

> *There's still some situations that catch me off guard. You know, some people saying things that I've not come across before, experiencing things that I've not come across before.* [MV5]

Behavioral Regulation (Psychological Capability): Holding Back on Advice

Having a natural inclination to offer advice made it difficult for some moderators to self-monitor their responses. Going against instincts and refraining from giving advice when young people shared their problems required self-monitoring of behavior. One moderator expressed the following:

> *It is like to go against that nature and leave it open to the room to support each other and not give advice.* [MS1]

As moderators found it difficult to hold back, effective moderation became challenging. One moderator stated the following:

> *We can't give medical advice, it's tough...we shouldn't really be saying (giving advice) to someone.* [MV5]

Belief About Consequences (Reflective Motivation): Impact of Moderation

Overall, the perceived outcome of moderation for young people, as understood by moderators, facilitated their moderation. Moderators had different beliefs on the impact that moderation had on young people. Some believed that moderation helped create a safe space. One moderator expressed the following:

> *I think young people find it really reassuring to have, like, moderators there because there's a lot of online support spaces, but they're not always moderated or kept safe.* [MS2]

Some moderators felt that young people could feel left out and unsupported. One moderator stated the following:

> *And sometimes the young people can feel like the moderators are not responding to them.* [MS3]

The vast coverage of the impact of moderation was also highlighted by moderators. One moderator expressed the following:

> *I think you do notice the impact because it covers everything OK, like mental health is the main domain. But it's also like everything like education, careers, drugs and alcohol. Just like everything that someone could be going through. So I think you do notice a wider impact.* [MV4]

Social or Professional Role and Identity (Reflective Motivation): Congruence Between Social and Professional Identity

The alignment of personal values and social identity with the professional role of a moderator facilitated moderation for all moderators. One moderator expressed the following:

> *I love volunteering and I love working in a church in the voluntary sector and working with charities. So and I like working with young people. A lot of my work has been with young people. So and yeah, I think it does. It does fit in that respect personally.* [MS3]

The role of a moderator aligned with the identity that moderators had built for themselves. One moderator stated the following:

> *I also work in mental health in my 9 to five job, so that really helps and that's why I wanted to start it in the first place because I started the chat at university and I was trying to broaden my experience working with people with mental health conditions.* [MS7]

Intentions (Reflective Motivation): Intending to Shape Young People's Perceptions

Some moderators expressed a deliberate intention to shape and manage young people's perceptions and expectations of the chat. One moderator expressed the following:

> *I've tried over the years to change the young peoples' perception of what the chat should be used for, I think that in the spirit of trying to make it more of a group conversation like a group chat.* [MS1]

This intention served as a strong facilitator for moderators. One moderator stated the following:

> *So I guess how we manage those situations, yeah, we can warn community members like, we encourage them to take a step back and let the moderators manage the situation. And yeah, we always let them know like they might be removed from the room if they don't like, listen to us.* [MS2]

Emotion (Automatic Motivation): Distressing Subject Matter

Most moderators identified that topics discussed in group chats, such as self-harm and suicidal thoughts, could take an emotional toll on moderators and cause stress and exhaustion. One moderator stated the following:

> *The topics that are being discussed are quite distressing. Examples are feeling very, very low and sometimes it becomes a bit of an echo chamber of you know, they're all talking about feeling like they want to end their life or feeling like self-harming.* [MS1]

These topics also gave way to discussions about the larger mental health ecosystem and evoked concern among the moderators for the young people. One moderator expressed the following:

> *Hard when you see them struggling. And so yeah, I think when that happens, you like worry quite a lot and it does kind of be on your mind for like a few days.* [MS2]

These distressing thoughts and related feelings inevitably acted as a barrier for moderators.

Environmental Context and Resources (Physical Opportunity)

Organizational Resources

The multitude of resources provided by The Mix was appreciated by all moderators. These encompassed guidelines, debrief forms, newsletters, and more, helping them moderate effectively. One moderator stated the following:

> *It really helps to have that handbook there to walk you through those guidelines and how to respond.* [MS3]

Moderation was facilitated through the ready availability of these resources, providing moderators with content that could be used in the chat. One moderator expressed the following:

> *And we have our Mod handbook which is really helpful in terms of giving us like example messages to pop in the group chat and to manage certain situations.* [MS7]

Lack of Visual Cues

Working in a web-based environment meant that moderators had to work without any visual cues, such as facial expressions. One moderator stated the following:

> *I think moderating is quite a unique task...especially when it's online.[MS2]*

This absence made it difficult for moderators to gauge the intensity of young people's emotions and thus required extra effort and attention, which could be a barrier for some moderators. One moderator expressed the following:

> *I guess you have to put a bit of extra effort into that because you don't have the visual cues of like, nodding and all that stuff.* [MS1]

Social Influences (Social Opportunity): Support Offered

Support offered during and after moderation (eg, support from supervisors and emails checking up on moderators) acted as an enabling influence. All respondents had an overwhelmingly positive response toward the support that they received from their supervisors and believed that this support helped them moderate. A respondent stated the following:

> *[The Mix] never makes you feel unappreciated. Like we get quite regular emails of just a reminder that you're all doing really good things...and I think yeah, when you have had a difficult chat and sort of things, they're feeling a bit hard, just things like that really make a difference.* [MS7]

In addition, volunteers reported that the presence of a supervisor in the chats made moderation easier. One moderator stated the following:

> *The moderator would flag it to the supervisor on shift and just say you know ohh it looks like these two are having a bit of a disagreement. What should we do?* [MS1]

One moderator expressed the following:

> *They are really understanding about wanting someone else to step in with a conversation that hits too close to home or needing 5 mins to yourself during the session. There is also a debrief form and if I say anything I felt unhappy with, they always chase it up to check on me or ask if I want more training in the area.* [MV6]

BCTs to Address Barriers

To optimize user engagement, Table 4 presents the BCTs identified to address the core barriers and enhance the facilitators, along with examples.

Table 5 shows the potential intervention types along with the corresponding BCTs to overcome the identified barriers to moderation.

Table 4. Barriers to engagement and strategies for change.

TDF[a] domain	Barrier	Intervention types	Potential BCTs[b]	Operationalization of BCTs
Cognitive and interpersonal skills	Difficulty expressing feelings and needs	Training	Instruction on how to perform a behavior (4.1[c]); behavioral practice/rehearsal (8.1)	Self-reflection worksheets: provide optional worksheets with guidance on identifying emotions and the support needed before users' participation in the group chat.
Memory, attention, and decision processes	Memory recall of guidelines	Enablement	Commitment (1.9)	Group agreement (in the support chat): collectively agree to abide by principles and behaviors aligned with the guidelines at the start of the support chat.
Memory, attention, and decision processes	Memory recall of guidelines	Education	Prompts/cues (7.1)	Reminders: implement periodic reminders or prompts about the chat guidelines to reinforce their importance and improve memory recall.
Beliefs about consequences	Judgment and confidentiality concerns	Education	Information about emotional consequences (5.6)	Share success stories: highlight success stories or testimonials from other young individuals who have benefited from engaging in peer-to-peer group chats.
Emotions	Fear and anxiety	Enablement	Reduce negative emotions (11.2)	Clarify burden misconceptions: provide information and resources from peers about the value of sharing and supporting each other, and share examples of challenges discussed by previous service-users in groups chats to address misconceptions that sharing is a burden and facilitate engagement.
Environmental context and resources	Overload and lack of structure	Environmental restructuring	Restructuring the physical environment (12.1)	Group similar topics and establish topic rotation: encourage moderators and members to brainstorm and group similar topics together either during the support chat (eg, via a poll website such as Slido) or before the chat (eg, via a poll on the discussion boards), allowing for more focused and meaningful discussions instead of fragmented conversations (alternative: create more themed chat sessions).
Environmental context and resources	Overload and lack of structure	Environmental restructuring	Restructuring the social environment (12.2)	Limit chat size: implement a cap on the number of members allowed in the support chat at 1 time to maintain a manageable and supportive group size.
Environmental context and resources	Overload and lack of structure	Environmental restructuring	Adding objects to the environment (12.5)	Emoticons: consider moving to a new chat software that enables members to offer immediate reactions to others' messages via emoticons rather than messages.
Social influences	Integration of new users and gaps in support	Enablement	Social support (unspecified; 3.1) and social support (practical; 3.2)	Buddy system: assign new users a designated "welcoming buddy" (peer) whose role is to encourage or practically facilitate interactions during their initial sessions; grouping or pairing system: pairing or grouping members in the chat as each other's dedicated "supporters" for the session to enhance peer support and encourage more engagement.

[a]TDF: Theoretical Domains Framework.

[b]BCT: behavior change technique.

[c]Behaviour Change Technique code numbers as per the Behaviour Change Technique Taxonomy provided in Michie et al [21].

Table 5. Barriers to moderation and strategies for change.

TDF[a] domain	Barrier	Intervention types	Potential BCTs[b]	Operationalization of BCTs
Memory, attention, and decision processes	Responding quickly to complex situations	Training	Behavioral practice/rehearsal (8.1[c])	Practice sessions: create practice moderation sessions that simulate the quick nature of group chats, which could help moderators to enhance their attention and decision-making processes.
Behavioral regulation	Holding back on advice	Training	Self-monitoring of behavior (2.3)	Diaries: ask moderators to make a note of situations and monitor when they feel instinctively inclined to offer advice to avoid it in the future.
Emotion	Distressing subject matter	Enablement	Reduce negative emotions (11.2)	Positive messaging: providing information about the positive impact of moderation through regular updates (eg, weekly or biweekly emails such as "Here's the impact that you helped deliver!") to increase moderators' positive emotions and reduce concerns about young people.
Environmental context and resources	Lack of visual cues	Environmental restructuring	Prompts/cues (7.1)	Phrase book: prompts or cues may be provided to compensate for the lack of visual cues. A booklet of phrases or "words to look out for" may be provided such that moderators may look out for these words to gauge if an individual is at risk, which may otherwise be missed.

[a]TDF: Theoretical Domains Framework.

[b]BCT: behavior change technique.

[c]Behaviour Change Technique code numbers as per the Behaviour Change Technique Taxonomy provided in Michie et al [21].

Discussion

Principal Findings

This study addresses a research gap by investigating the influences on engagement with, and moderation of, synchronous web-based peer support group chats to support young people's mental health. Using the BCW framework, thematic analysis revealed a total of 20 themes, 9 (45%) for engagement and 11 (55%) for moderation. Of the 20 themes, 3 (15%) were facilitators of engagement, 7 (35%) were facilitators of moderation, 4 (20%) were barriers to moderation, and 6 (30%) were barriers to engagement. The following discussion focuses on the COM-B and TDF themes that were prominent and common across both user engagement and moderation. The findings of this study are discussed in relation to previous research, and the potential BCTs to address the identified barriers to engagement and moderation are contextualized.

Facilitators Common to Moderation and Engagement

Cognitive and interpersonal skills and knowledge (psychological capability) enabled users to express their feelings and needs. Both moderators and users benefitted from having the skills to listen and respond to other users. Moderators developed these skills through practice and were supported by a strong knowledge and understanding of the chat guidelines. Similar themes have been found in other modes of web-based mental health support [39].

Intentions (reflective motivation) were also core themes for moderators and users—both were driven by their desire to help others and ensure that the chat was a valuable resource for users. This sentiment is echoed elsewhere in the literature on web-based support groups, where listeners aimed to create a "safe and warm" space for their clients [51]. Similarly, moderators in this study consciously resolved to provide a safe and nonjudgmental space for young people. This is a theme that can be seen elsewhere in the literature, relating to other web-based peer mental health communities [30].

Finally, for both engagement and moderation, social influences (social opportunity) was another facilitator, where other users, moderators, or supervisors were seen as understanding, supportive, and appreciative of others' needs. For users, this meant being able to share experiences in a safe environment, and moderators felt valued, appreciated, and encouraged in their moderation. Such support also promotes moderators' mental health. As has been previously suggested by Aldamman et al [52], perceived organizational support was positively related to mental well-being, reduced emotional exhaustion, and reduced stress among humanitarian volunteers. This aspect is also linked to the potential of moderators to contribute to the enhancement of the mental well-being of service users, as demonstrated by Perry et al [19].

Barriers Common to Moderation and Engagement

Group chats present a complex environment due to their dynamic and fast-paced nature, with multiple users. Users could find it difficult to remember and adhere to the guidelines, especially in pressurized situations (eg, crises). Although moderators were familiar with the guidelines, they found it challenging to make real-time decisions, particularly during busy chats when many users were interacting rapidly with one another. This inherent complexity can function as a barrier to effective moderation and engagement. With a continuous inflow of new messages in group chats, it becomes challenging to focus on messages and identify any specific theme, as noted by Li et al [53]. Moderators in this study also encountered difficulties in promptly and accurately assessing messages for potential at-risk situations, under time pressure, while simultaneously

ensuring that all the young people in the chat felt supported. The BCT of "Behavioral practice" can help equip moderators to handle novel situations quickly as they arise. This BCT has been successfully used to "positively influence" behavior in the context of user training [54] and in interventions supporting shared decision-making [55].

The challenges with rapid decision-making of how to respond partly resulted from the chat environment being unstructured and uncontrolled in terms of the volume and speed of message exchanges; this was a barrier for moderators and users. Implementing emoticons (BCT "adding objects to the environment") could offer an alternative and rapid way of engaging with messages, helping to express emotions while reducing message overload [56]. Previous studies have also highlighted participants of a Cognitive Behavioral Therapy-based peer support platform feeling overwhelmed and stressed due to message volume and unfamiliar dynamics [57]. This underscores the need for accessible and well-organized chat platforms. One promising BCT is "restructuring the physical environment." Similar topics could be grouped within the platform or support chat sessions could be themed to allow users to explore and self-organize into groups based on shared characteristics and pain points, as has been implemented elsewhere [58].

Another aspect of the environment that was highlighted in previous studies was the lack of visual clues to support decision-making. This was a particular barrier for moderators, impeding connection and relationship building [39,59]. The BCT "prompts/cues" has been previously suggested in a similar context to improve moderation [39]. These could be in the form of electronic prompts or suggested phrases that pop-up when a young person writes a phrase that may need to be flagged. Such prompts may support moderators to identify individuals at risk and compensate for the lack of visual cues. This BCT has been successfully used to prompt action by the user in various contexts [60] and can also be used to simultaneously address barriers related to memory difficulties.

Given the sensitive nature of the discussions within web-based peer support chats, it is perhaps unsurprising that emotions can constitute another barrier. The BCT of "reduce negative emotions" can help address this. For users, the fear and anxiety associated with disclosure can prevent them from sharing in the chats, hindering their access to the help they need. Previous interventions report participants finding it helpful to know that others were undergoing similar emotions, reducing feelings of isolation [61]. Therefore, providing information and resources from peers about the value of sharing and supporting each other, and sharing examples of challenges discussed in group chats by previous service users may facilitate engagement. This could further build a sense of a supportive and friendly community on the platform, which facilitates willingness to share feelings and difficulties in other contexts [57]. Dealing with distressing topics such as mental health issues, suicidal thoughts, and other issues that young people are dealing with often becomes emotionally exhausting for moderators. In this case, the BCT "reduce negative emotions" could involve providing moderators with information about the positive impact of their moderation, to counter any negative emotions. A scoping review highlighted that this BCT has frequently been used in the development of mental health interventions [62]. Resilience-building training may also be helpful in equipping moderators to deal with the "emotional cost" of content moderation [63].

Barriers Specific to User Engagement

The barrier of "difficulties expressing feelings and support needs" experienced by users aligns with a prior study on webchat counseling engagement among young people, which found users lacking self-expression skills [28]. "Constructive emotional sharing" is a skill that can be improved with practice [64]. Promising BCTs to address this barrier include "instruction on how to perform a behavior" and "behavioral practice/rehearsal." For example, self-reflection worksheets could encourage users to identify their feelings and practice written statements before participating in group chats [64]. Statements such as "I feel...when...because" could be shared, leading to more personalized responses [65]. These BCTs could prove particularly beneficial for newcomers or individuals less familiar with the dynamics of group chat interaction, also creating positive spillover to a further barrier "integration of new users and gaps in support." A barrier identified here aligns with a previous study, which found that receiving empathic comments initially has a significant cascading effect, motivating individuals to reciprocate and offer support to others [66]. To further encourage integration and support for new users, the BCT "social support" could be introduced by assigning a designated "welcoming buddy" whose role is to support new members during their initial sessions or pairing members in the chat as each other's "supporter" to enhance peer support and encourage more engagement. Prior research has shown that shared interests encouraged conversation between new pairs in a peer support intervention, who were strangers when they were initially paired [67].

Having judgment and confidentiality concerns constituted another barrier. Overcoming such concerns can contribute to more fluid communication in web-based peer support and are a core component of building trust [67]. The BCT "information about emotional consequences" may address this barrier through sharing success stories, statistics, or personal narratives. Within the recovery model, this increases a sense of acceptance, understanding, and authenticity, particularly for new service users who tend to experience these fears more than long-term users [68].

Barriers Specific to Moderation

Moderators found it difficult to hold back on offering advice. Rooted in the logic of care [69], moderation is contextualized and involves an empathetic dialogue or interaction between moderators and users [70]. In such an open-ended process, holding back, given the natural flow of an interaction, proved a challenge for moderators. "Self-monitoring of behavior" could be a potential BCT that may help moderators to keep their behavior in check and change it. This could be achieved by encouraging moderators to note when they instinctively offer advice or feel like they want to do so, allowing them to reflect and monitor their own behavior if such situations arise again. This technique has proved effective in many behavior change

interventions in different contexts [71,72], particularly for health behaviors [73,74].

Limitations

The findings of this study need to be considered in light of certain limitations. The sampling of this study was reliant on self-selection, which may have introduced bias and favored participation by more confident users and moderators, potentially influencing the barriers and facilitators identified [75]. Although the participants were assured that their responses would remain confidential and would not have any consequences with respect to their relationship with The Mix, they may nevertheless have given socially desirable responses, whether intentionally or subconsciously [76]. Furthermore, it is noteworthy that the study participants were predominately women, making the sample relatively homogeneous. However, the overall population of moderators and group chat users at The Mix is also mainly women (>70% of users and 95% of moderators). Finally, the study's findings are grounded in the data obtained from a single mental health group chat forum, which could limit the generalizability of the results to other chat contexts operating under different circumstances.

Conclusions

Group chats are an increasingly popular form of digital mental health intervention, and this study contributes to building the evidence base, which can help optimize them as a safe and timely form of mental health support for young people. It is particularly valuable as it examines synchronous group chats, which are characterized by in-the-moment empathic interactions and emotional connections. Through using the COM-B and TDF, the study found that skills and knowledge, beliefs about consequences and intentions, emotions, and the social and physical environment are important factors influencing both the users and moderators of group chats. In particular, supporting the improvement of memory, attention, and decision-making skills of those involved; adapting the physical environment to facilitate effective interactions; and reducing negative emotions are suggested to optimize the value and effectiveness of group chats for young people's mental health support for both the users and moderators of these services. The intervention types and BCTs proposed serve to emphasize the importance of training and support, particularly for moderators, in this important role. The study also further demonstrates the effectiveness of the BCW approach and the use of the TDF and COM-B to understand the influences on behavior in a systematic manner, especially for mental health and well-being interventions.

A natural progression of this work would be to implement and evaluate the interventions proposed in the study and gauge to what extent the suggested BCTs reduce the identified barriers. In addition, the fidelity of the BCTs could also be assessed to understand the nuances of intervention delivery [77] to facilitate contextualized tailoring of the intervention. In an environment where digital mental health interventions for young people continue to grow in significance, this study and future suggested studies can contribute toward ensuring that they are evidence based and consider the voices of young people themselves.

Conflicts of Interest

None declared.

References

1. Ogden T, Hagen KA. Adolescent Mental Health: Prevention and Intervention. London, UK: Routledge; 2018.
2. Kauhanen L, Wan Mohd Yunus WM, Lempinen L, Peltonen K, Gyllenberg D, Mishina K, et al. A systematic review of the mental health changes of children and young people before and during the COVID-19 pandemic. Eur Child Adolesc Psychiatry 2023 Jun 12;32(6):995-1013 [FREE Full text] [doi: 10.1007/s00787-022-02060-0] [Medline: 35962147]
3. Creswell C, Shum A, Pearcey S, Skripkauskaite S, Patalay P, Waite P. Young people's mental health during the COVID-19 pandemic. The Lancet Child Adolescent Health 2021 Aug;5(8):535-537. [doi: 10.1016/s2352-4642(21)00177-2]
4. Mental Health of Children and Young People in England 2022 - wave 3 follow up to the 2017 survey. National Health Service England. 2022 Nov 29. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2022-follow-up-to-the-2017-survey [accessed 2025-01-12]
5. Mental Health of Children and Young People in England, 2023 - wave 4 follow up to the 2017 survey. National Health Service England. 2023 Nov 21. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2023-wave-4-follow-up [accessed 2025-01-12]
6. McGorry PD, Mei C, Chanen A, Hodges C, Alvarez-Jimenez M, Killackey E. Designing and scaling up integrated youth mental health care. World Psychiatry 2022 Feb 11;21(1):61-76 [FREE Full text] [doi: 10.1002/wps.20938] [Medline: 35015367]
7. Facts and figures. Children & Young People's Mental Health Coalition. URL: https://cypmhc.org.uk/resources/facts-and-figures/ [accessed 2025-01-12]
8. Yang D, Kraut RE, Smith T, Mayfield E, Jurafsky D. Seekers, providers, welcomers, and storytellers: modeling social roles in online health communities. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK. [doi: 10.1145/3290605.3300574]
9. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015 May 19;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]

10. Castelnuovo G, Gaggioli A, Mantovani F, Riva G. From psychotherapy to e-therapy: the integration of traditional techniques and new communication tools in clinical settings. Cyberpsychol Behav 2003 Aug;6(4):375-382. [doi: 10.1089/109493103322278754] [Medline: 14511449]
11. Dowling M, Rickwood D. Online counseling and therapy for mental health problems: a systematic review of individual synchronous interventions using chat. J Technol Hum Serv 2013 Jan;31(1):1-21 [FREE Full text] [doi: 10.1080/15228835.2012.728508]
12. Tibbs M, O'Reilly A, Dwan O'Reilly M, Fitzgerald A. Online synchronous chat counselling for young people aged 12-25: a mixed methods systematic review protocol. BMJ Open 2022 Apr 25;12(4):e061084 [FREE Full text] [doi: 10.1136/bmjopen-2022-061084] [Medline: 35470202]
13. Naslund JA, Deng D. Addressing mental health stigma in low-income and middle-income countries: a new frontier for digital mental health. Ethics Med Public Health 2021 Dec;19:100719 [FREE Full text] [doi: 10.1016/j.jemep.2021.100719] [Medline: 35083375]
14. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
15. Griffiths KM, Reynolds J, Vassallo S. An online, moderated peer-to-peer support bulletin board for depression: user-perceived advantages and disadvantages. JMIR Ment Health 2015 Apr 24;2(2):e14 [FREE Full text] [doi: 10.2196/mental.4266] [Medline: 26543919]
16. Easton K, Diggle J, Ruethi-Davis M, Holmes M, Byron-Parker D, Nuttall J, et al. Qualitative exploration of the potential for adverse events when using an online peer support network for mental health: cross-sectional survey. JMIR Ment Health 2017 Oct 30;4(4):e49 [FREE Full text] [doi: 10.2196/mental.8168] [Medline: 29084710]
17. Merchant R, Goldin A, Manjanatha D, Harter C, Chandler J, Lipp A, et al. Opportunities to expand access to mental health services: a case for the role of online peer support communities. Psychiatr Q 2022 Jun 28;93(2):613-625 [FREE Full text] [doi: 10.1007/s11126-022-09974-7] [Medline: 35226251]
18. Wadden D, August T, Li Q, Althoff T. The effect of moderation on online mental health conversations. Proc Int AAAI Conf Web Soc Media 2021;15(1):751-763. [doi: 10.1609/icwsm.v15i1.18100]
19. Perry A, Pyle D, Lamont-Mills A, du Plessis C, du Preez J. Suicidal behaviours and moderator support in online health communities: a scoping review. BMJ Open 2021 Jun 30;11(6):e047905 [FREE Full text] [doi: 10.1136/bmjopen-2020-047905] [Medline: 34193497]
20. Kraut RE, Resnick P, Kiesler S. Building Successful Online Communities: Evidence-Based Social Design. Cambridge, MA: MIT Press; 2012.
21. Michie S, Atkins L, West R. The Behaviour Change Wheel: A Guide to Designing Interventions. Sutton, UK: Silverback Publishing; 2014.
22. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun 13;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
23. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
24. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
25. Lakind D, Bradley WJ, Patel A, Chorpita BF, Becker KD. A multidimensional examination of the measurement of treatment engagement: implications for children's mental health services and research. J Clin Child Adolesc Psychol 2022 Jul 16;51(4):453-468 [FREE Full text] [doi: 10.1080/15374416.2021.1941057] [Medline: 34269632]
26. James TL, Villacis Calderon ED, Bélanger F, Lowry PB. The mediating role of group dynamics in shaping received social support from active and passive use in online health communities. Inf Manag 2022 Apr;59(3):103606. [doi: 10.1016/j.im.2022.103606]
27. Liverpool S, Mota CP, Sales CM, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020 Jun 23;22(6):e16317 [FREE Full text] [doi: 10.2196/16317] [Medline: 32442160]
28. Mawdsley G, Richiello M, Gutman LM. Barriers and facilitators of young people's engagement with webchat counselling: a qualitative analysis informed by the Behaviour Change Wheel. Couns Psychother Res 2022 Mar 23;22(3):725-735. [doi: 10.1002/capr.12528]
29. Onken SJ, Dumont JM, Ridgway P, Dornan DH, Ralph RO. Mental health recovery: what helps and what hinders? A national research project for the development of recovery facilitating system performance indicators. National Association of State Mental Health Program Director and National Technical Assistance Center for State Mental Health Planning. 2002 Oct. URL: https://www.nasmhpd.org/sites/default/files/MHSIPReport%281%29.pdf [accessed 2025-01-12]

30. Deng D, Rogers T, Naslund JA. The role of moderators in facilitating and encouraging peer-to-peer support in an online mental health community: a qualitative exploratory study. J Technol Behav Sci 2023 Feb 16;8(2):128-139 [FREE Full text] [doi: 10.1007/s41347-023-00302-9] [Medline: 36810998]
31. Prescott J, Rathbone AL, Brown G. Online peer to peer support: qualitative analysis of UK and US open mental health Facebook groups. Digit Health 2020 Dec 10;6:2055207620979209 [FREE Full text] [doi: 10.1177/2055207620979209] [Medline: 33354335]
32. Strand M, Eng LS, Gammon D. Combining online and offline peer support groups in community mental health care settings: a qualitative study of service users' experiences. Int J Ment Health Syst 2020 May 29;14(1):39 [FREE Full text] [doi: 10.1186/s13033-020-00370-x] [Medline: 32514303]
33. Saha K, Ernala SK, Dutta S, Sharma E, de Choudhury M. Understanding moderation in online mental health communities. In: Proceedings of the 12th International Conference on Social Computing and Social Media. Participation, User Experience, Consumer Experience, and Applications of Social Computing. 2020 Presented at: SCSM 2020; July 19-24, 2020; Copenhagen, Denmark. [doi: 10.1007/978-3-030-49576-3_7]
34. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb 01;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
35. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7(1):37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
36. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
37. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
38. Richiello MG, Mawdsley G, Gutman LM. Using the Behaviour Change Wheel to identify barriers and enablers to the delivery of webchat counselling for young people. Couns Psychother Res 2021 May 04;22(1):130-139. [doi: 10.1002/capr.12410]
39. Perowne R, Gutman LM. Barriers and enablers to the moderation of self-harm content for a young person's online forum. J Ment Health 2024 Jun 16;33(3):357-365 [FREE Full text] [doi: 10.1080/09638237.2022.2069721] [Medline: 35574666]
40. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
41. Azungah T. Qualitative research: deductive and inductive approaches to data analysis. Qual Res J 2018 Oct 31;18(4):383-400. [doi: 10.1108/qrj-d-18-00035]
42. Bradley EH, Curry LA, Devers KJ. Qualitative data analysis for health services research: developing taxonomy, themes, and theory. Health Serv Res 2007 Aug 24;42(4):1758-1772 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00684.x] [Medline: 17286625]
43. Maguire M, Delahunt B. Doing a thematic analysis: a practical, step-by-step guide for learning and teaching scholars. All Ireland J High Educ 2017 Oct 31;9(3) [FREE Full text] [doi: 10.62707/aishej.v9i3.335]
44. Hallberg LR. The “core category” of grounded theory: making constant comparisons. Int J Qual Stud Health Well Being 2009 Jul 12;1(3):141-148. [doi: 10.1080/17482620600858399]
45. Berger R. Now I see it, now I don’t: researcher’s position and reflexivity in qualitative research. Qual Res 2013 Jan 03;15(2):219-234. [doi: 10.1177/1468794112468475]
46. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. Br J Health Psychol 2015 Feb 12;20(1):130-150. [doi: 10.1111/bjhp.12102] [Medline: 24815766]
47. Johnston M, Carey RN, Connell Bohlen LE, Johnston DW, Rothman AJ, de Bruin M, et al. Development of an online tool for linking behavior change techniques and mechanisms of action based on triangulation of findings from literature synthesis and expert consensus. Transl Behav Med 2021 May 25;11(5):1049-1065 [FREE Full text] [doi: 10.1093/tbm/ibaa050] [Medline: 32749460]
48. Welcome to the Theory and Techniques Tool. Theory and Techniques Tool. URL: https://theoryandtechniquetool.humanbehaviourchange.org/ [accessed 2025-01-12]
49. Allison AL, Purkiss D, Lorencatto F, Miodownik M, Michie S. Improving compostable plastic disposal: an application of the Behaviour Change Wheel intervention development method. Front Sustain 2022 Aug 18;3 [FREE Full text] [doi: 10.3389/frsus.2022.968152]
50. Gutman LM, Perowne R, Younas F, O'Hanrachtaigh E. Making hybrid work for diverse staff in higher education: a behaviour change approach. High Educ Q 2023 Dec 15;78(3):784-806. [doi: 10.1111/hequ.12487]
51. Barak A. Emotional support and suicide prevention through the internet: a field project report. Comput Hum Behav 2007 Mar;23(2):971-984. [doi: 10.1016/j.chb.2005.08.001]

52. Aldamman K, Tamrakar T, Dinesen C, Wiedemann N, Murphy J, Hansen M, et al. Caring for the mental health of humanitarian volunteers in traumatic contexts: the importance of organisational support. Eur J Psychotraumatol 2019 Dec 03;10(1):1694811 [FREE Full text] [doi: 10.1080/20008198.2019.1694811] [Medline: 31839900]
53. Li H, Kraut RE, Zhu H. Technical features of asynchronous and synchronous community platforms and their effects on community cohesion: a comparative study of forum-based and chat-based online mental health communities. J Comput Mediat Commun 2021;26(6):403-421. [doi: 10.1093/jcmc/zmab016]
54. Wood CE, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. Applying the behaviour change technique (BCT) taxonomy v1: a study of coder training. Transl Behav Med 2015 Jun 19;5(2):134-148 [FREE Full text] [doi: 10.1007/s13142-014-0290-z] [Medline: 26029276]
55. Agbadjé TT, Elidor H, Perin MS, Adekpedjou R, Légaré F. Towards a taxonomy of behavior change techniques for promoting shared decision making. Implement Sci 2020 Aug 20;15(1):67 [FREE Full text] [doi: 10.1186/s13012-020-01015-w] [Medline: 32819410]
56. Aldunate N, Villena-González M, Rojas-Thomas F, López V, Bosman CA. Mood detection in ambiguous messages: the interaction between text and emoticons. Front Psychol 2018 Apr 4;9:423 [FREE Full text] [doi: 10.3389/fpsyg.2018.00423] [Medline: 29670554]
57. Browne NL, Carragher NO, O'Toole A, Pimm J, Ryder J, Thew GR. Evaluating user experiences of SHaRON: an online CBT-based peer support platform. The Cogn Behav Ther 2022 Apr 11;15:e18. [doi: 10.1017/s1754470x22000150]
58. O'Leary K, Bhattacharya A, Munson SA, Wobbrock JO, Pratt W. Design opportunities for mental health peer support technologies. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR. [doi: 10.1145/2998181.2998349]
59. Lau PL, Jaladin RA, Abdullah HS. Understanding the two sides of online counseling and their ethical and legal ramifications. Procedia Soc Behav Sci 2013 Nov;103:1243-1251. [doi: 10.1016/j.sbspro.2013.10.453]
60. Morrissey EC, Corbett TK, Walsh JC, Molloy GJ. Behavior change techniques in apps for medication adherence: a content analysis. Am J Prev Med 2016 May;50(5):e143-e146. [doi: 10.1016/j.amepre.2015.09.034] [Medline: 26597504]
61. Borghouts J, Eikey E, Mark G, de Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
62. Martinengo L, Jabir AI, Goh WWT, Lo NYW, Ho MHR, Kowatsch T, et al. Conversational agents in health care: scoping review of their behavior change techniques and underpinning theory. J Med Internet Res 2022 Oct 03;24(10):e39243 [FREE Full text] [doi: 10.2196/39243] [Medline: 36190749]
63. Steiger M, Bharucha TJ, Venkatagiri S, Riedl MJ, Lease M. The psychological well-being of content moderators: the emotional labor of commercial moderation and avenues for improving support. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445092]
64. van Dyke K. Feeling stuck? How to express your feelings. Psych Central. 2021 Jun 17. URL: https://psychcentral.com/lib/feeling-stuck-how-to-express-your-feelings [accessed 2025-01-12]
65. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX. [doi: 10.1145/2207676.2208602]
66. Chen Y, Xu Y. Exploring the effect of social support and empathy on user engagement in online mental health communities. Int J Environ Res Public Health 2021 Jun 26;18(13):6855 [FREE Full text] [doi: 10.3390/ijerph18136855] [Medline: 34206719]
67. Andalibi N, Flood MK. Considerations in designing digital peer support for mental health: interview study among users of a digital support system (Buddy Project). JMIR Ment Health 2021 Jan 04;8(1):e21819 [FREE Full text] [doi: 10.2196/21819] [Medline: 33393909]
68. Peck CE, Lim MH, Purkiss M, Foley F, Hopkins L, Thomas N. Development of a lived experience-based digital resource for a digitally-assisted peer support program for young people experiencing psychosis. Front Psychiatry 2020 Jul 2;11:635 [FREE Full text] [doi: 10.3389/fpsyt.2020.00635] [Medline: 32714223]
69. Mol A. The Logic of Care: Health and the Problem of Patient Choice. Milton Park, UK: Taylor & Francis; 2008.
70. Ruckenstein M, Turunen LL. Re-humanizing the platform: content moderators and the logic of care. New Media Soc 2019 Sep 19;22(6):1026-1042. [doi: 10.1177/1461444819875990]
71. Keller J, Kwasnicka D, Wilhelm LO, Lorbeer N, Pauly T, Domke A, et al. Hand washing and related cognitions following a brief behavior change intervention during the COVID-19 pandemic: a pre-post analysis. Int J Behav Med 2022 Oct 29;29(5):575-586 [FREE Full text] [doi: 10.1007/s12529-021-10042-w] [Medline: 34843096]
72. Berry R, Kassavou A, Sutton S. Does self-monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta-analysis. Obes Rev 2021 Oct 30;22(10):e13306. [doi: 10.1111/obr.13306] [Medline: 34192411]
73. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]

74. Garnett CV, Crane D, Brown J, Kaner EF, Beyer FR, Muirhead CR, et al. Behavior change techniques used in digital behavior change interventions to reduce excessive alcohol consumption: a meta-regression. Ann Behav Med 2018 May 18;52(6):530-543 [FREE Full text] [doi: 10.1093/abm/kax029] [Medline: 29788261]
75. Gautier PA, Klaauw BV. Selection in a field experiment with voluntary participation. J Appl Econometri 2012 Jan 18;27(1):63-84. [doi: 10.1002/jae.1184]
76. Bispo Júnior JP. Social desirability bias in qualitative health research. Rev Saude Publica 2022 Nov 18;56:101 [FREE Full text] [doi: 10.11606/s1518-8787.2022056004164] [Medline: 36515303]
77. Toomey E, Hardeman W, Hankonen N, Byrne M, McSharry J, Matvienko-Sikar K, et al. Focusing on fidelity: narrative review and recommendations for improving intervention fidelity within trials of health behaviour change interventions. Health Psychol Behav Med 2020 Mar 12;8(1):132-151 [FREE Full text] [doi: 10.1080/21642850.2020.1738935] [Medline: 34040865]

Abbreviations

BCT: behavior change technique
BCTT: Behavior Change Techniques Taxonomy
BCW: Behavior Change Wheel
COM-B: Capability, Opportunity, Motivation, and Behavior
TDF: Theoretical Domains Framework

Edited by A Kushniruk; submitted 08.07.24; peer-reviewed by D Levine, L Valentine; comments to author 24.11.24; revised version received 05.12.24; accepted 31.12.24; published 05.02.25.

Please cite as:
Ananya A, Tuuli J, Perowne R, Gutman LM
Barriers and Facilitators to User Engagement and Moderation for Web-Based Peer Support Among Young People: Qualitative Study Using the Behavior Change Wheel Framework
JMIR Hum Factors 2025;12:e64097
URL: https://humanfactors.jmir.org/2025/1/e64097
doi:10.2196/64097
PMID:

Original Paper

The Safety of Digital Mental Health Interventions: Findings and Recommendations From a Qualitative Study Exploring Users' Experiences, Concerns, and Suggestions

Rayan Taher[1], MSc; Daniel Stahl[2], PhD; Sukhi Shergill[1,3], MD, PhD; Jenny Yiend[1], PhD

[1]Department of Psychosis Studies, King's College London, London, United Kingdom

[2]Department of Biostatistics and Health Informatics, King's College London, London, United Kingdom

[3]Kent and Medway Medical School, Canterbury, United Kingdom

Corresponding Author:
Jenny Yiend, PhD
Department of Psychosis Studies
King's College London
Institute of Psychiatry, Psychology & Neuroscience (IoPPN)
16 De Crespigny Park
London, SE5 8AB
United Kingdom
Phone: 44 020 7848 0002
Email: jenny.yiend@kcl.ac.uk

Abstract

Background: The literature around the safety of digital mental health interventions (DMHIs) is growing. However, the user/patient perspective is still absent from it. Understanding the user/patient perspective can ensure that professionals address issues that are significant to users/patients and help direct future research in the field.

Objective: This qualitative study aims to explore DMHI users' experiences, views, concerns, and suggestions regarding the safety of DMHIs.

Methods: We included individuals aged 18 years old or older, having experience in using a DMHI, and can speak and understand English without the need for a translator. Fifteen individual interviews were conducted. Deductive thematic analysis was used to analyze the data.

Results: The analysis of the interview transcripts yielded 3 main themes: Nonresponse: A Concern, a Risk, and How Users Mitigate It, Symptom Deterioration and Its Management, and Concerns Around Data Privacy and How to Mitigate Them.

Conclusions: The results of this study led to 7 recommendations on how the safety of DMHIs can be improved: provide "easy access" versions of key information, use "approved by..." badges, anticipate and support deterioration, provide real-time feedback, acknowledge the lack of personalization, responsibly manage access, and provide genuine crisis support. These recommendations arose from users' experiences and suggestions. If implemented, these recommendations can improve the safety of DMHIs and enhance users' experience.

(JMIR Hum Factors 2025;12:e62974) doi:10.2196/62974

KEYWORDS

digital mental health; safety; user perspective; patient perspective; qualitative; risks; risk mitigation; deterioration; nonresponse; data safety

Introduction

Digital mental health interventions (DMHIs) are mental health interventions that are delivered through digital platforms such as mobile apps, websites, or virtual reality [1]. Some of the added benefits of mental health interventions that are delivered digitally are improved accessibility, scalability, convenience, and the potential for anonymous engagement [2]. To realize these benefits, users need to trust that these interventions are effective and safe [2]. The evidence shows that DMHIs can be as effective as traditional face-to-face therapies, especially for common mental health disorders such as depression and anxiety

[3,4]. However, the safety of DMHIs is still an evolving field [1,2,5,6].

DMHI's users face similar risks to those in face-to-face therapy, such as deterioration in symptoms, novel symptoms (experiencing new mental health symptoms during treatment), and nonresponse [1]. Deterioration of symptoms, observed in approximately 3%-10% of psychotherapy cases [7,8], signifies a phenomenon where patients' conditions worsen during therapy. Deterioration is the most common side effect of mental health therapies (face to face and digital) [1]. There is a debate in the literature about whether deterioration is a normal and integral part of therapy or an unnecessary side effect [9,10]. A recent experts' consensus study concluded that short-term deterioration that occurs during therapy is part of therapy and should not be considered a safety concern; however, deterioration still needs to be monitored to ensure that it is not chronic, severe, and does not lead to an adverse event such as the patient dropping out of treatment [11]. Nonresponse occurs when the therapy is not effective in relieving the target symptoms [12]. It is considered a negative outcome as it hinders access to more effective treatments, spontaneous remission, and may prolong or even increase distress [12]. Additionally, the digital nature of DMHIs introduces additional risks to mental health therapies such as technical issues and privacy concerns [5,13].

The expanding body of literature addressing the safety of DMHIs is noteworthy [1,5,6]. While considerable attention has been devoted to exploring the safety of DMHIs (how it is assessed, analyzed, and reported), a notable gap exists in the qualitative understanding of individual perspectives on the topic. Existing qualitative studies in this field have either focused on the viewpoints of health professionals and medical students [14,15] or have sought user opinions on specific digital innovations usually as part of a wider program of development work. A few studies have sought service users' views on digital interventions more generally [16-18], but no work to our knowledge has investigated user/patient perspectives specifically about the safety of these technologies [19,20]. Understanding the user/patient perspective on safety can help direct future research in the field to ensure professionals focus on issues significant to users/patients; identifying users' concerns can help professionals address these issues, leading to higher rates of adherence and engagement. For that reason, this qualitative study aims to explore DMHIs users' experiences, views, concerns, and suggestions regarding the safety of DMHIs.

Methods

Design and Aim

This qualitative study utilized individual interviews to explore users' experiences, views, concerns, and suggestions about the safety of DMHIs. The Consolidated Criteria for Reporting Qualitative Research (COREQ) was used to report the results of this study [21].

Recruitment

The study included individuals aged 18 years or older, with previous experience using a DMHI, and speaking and understanding English without the need for a translator. The DMHI needed to be a mental health intervention that was provided via a tech-based medium (eg, app, website, virtual reality) and targeted a specific mental health condition. Participants were recruited using nonpurposive sampling by posting advertisements on authors' own professional social media platforms, such as X (formerly Twitter; X Corp.) and LinkedIn (Microsoft Corporation), an online participant recruiting platform (MQ; MQ Mental Health Research), the university's (King's College London) research volunteering circular email, and through DMHI trials whose participants consented for their details to be shared for future research. A total of 54 potential participants reached out; 15 (28%) were eligible and participated in the study, 7 (13%) did not complete the online eligibility form, and 32 (59%) were ineligible. Eligibility was determined through screening questions asked of potential participants to gather further details about the specific intervention used. The main reasons for ineligibility were that the intervention used was telehealth (eg, face-to-face therapy conducted via video call) or did not target a specific mental health condition (eg, mindfulness apps).

Participants

A total of 15 participants were recruited to participate in this study. Researchers initially estimated a sample size of 6-16 participants based on pragmatic recommendations from the literature, suggesting that 6-16 interviews provide sufficient information power [22]. Then the final sample size (15 participants) was determined based on the richness of the data and their ability to sufficiently answer the research question [23]. Of the 15 participants, 12 (80%) were females and 3 (20%) were males. All participants lived in the United Kingdom at the time of the interviews. Participants had an average age of 30 years (SD 6.43 years; range 19-42 years). On average, participants had used a DMHI for 8 months (SD 10.51 months; range 1-36 months). The DMHIs used by participants in this study were Beating the Blues, Calm Harm, FREED-M, Happify, Molehill Mountain, Moodkit, Silver Cloud, Sleepio, STOP app, Woebot, Youper, and an online intervention for Bulimia. See Table 1 for more information on the interventions used by participants in this study. Some participants used the DMHI for both anxiety and depression. Other than that, no other participants used the DMHI for more than 1 target symptom or condition.

Table 1. The DMHIs[a] used by participants in this study (N=15).

DMHI	Values, n (%)	Duration (months) of use, mean (range)
DMHI's target symptom/condition		
Depression and anxiety	5 (33)	3.4 (1-6)
Depression	2 (13)	19.25 (2.5-36)
Self-harm	2 (13)	12.75 (1.5-24)
Paranoia	1 (7)	3 (N/A[b])
Insomnia	1 (7)	1.5 (N/A)
Postnatal depression	1 (7)	24 (N/A)
Bulimia	1 (7)	1.5 (N/A)
Anxiety in autism	1 (7)	3 (N/A)
Eating disorders	1 (7)	3 (N/A)
DMHI's format		
Web-based	7 (47)	2.5 (1-6)
App-based	6 (40)	11.5 (1.5-36)
Artificial intelligence chatbot	2 (13)	15 (6-24)
Therapist involvement		
Self-administered (users independently used the DMHI without any support)	12 (80)	9.33 (1.5-36)
Hybrid (users independently used the DMHI while receiving regular support)	3 (20)	1.66 (1-2)
How participants found the DMHI		
Health care professional	7 (47)	N/A
Social media	2 (13)	N/A
App store	2 (13)	N/A
DMHI research	2 (13)	N/A
Work (via human resources)	1 (7)	N/A
University website	1 (7)	N/A

[a]DMHI: digital mental health intervention.

[b]N/A: not applicable.

Materials

Individual interviews with participants were conducted and recorded online via Microsoft Teams (Microsoft Corporation). Interviews were semistructured. See Multimedia Appendix 1 for the topic guide. The interviewer (RT) used prompts to facilitate and guide the discussion. RT has experience conducting individual interviews for research purposes.

Procedure

Participants viewing the study advertisement were asked to email the researcher if they were interested. The researcher replied to introduce them to the study, share the participant information sheet (PIS), and request that they complete an online (Qualtrics) questionnaire to check eligibility. The same researcher contacted eligible participants to check that they had read the PIS, answered any questions that they had, and asked if they were interested in participating in the study. Those expressing a desire to participate were emailed an online form that includes a few questions (details below) and an e-consent form (using Qualtrics) to sign and a link to book a 1-hour slot for the interview. The online form asked for demographic details such as gender and age, details about the intervention (name and intended purpose), and how long they used the intervention for. All interviews were audio recorded. At the end of the interview, participants were sent a thank you email and a £20 (US $25) voucher as compensation for their time. Interviews lasted on average 40 minutes. Recordings were automatically transcribed by Microsoft Teams and were verified for accuracy by RT. Once transcription was complete all recordings were deleted.

Analysis

Thematic Analysis Process

The 15 transcripts were uploaded onto NVivo (Lumivero, LLC) for analysis [24]. We used deductive thematic analysis to analyze the data. Thematic analysis is a process used to identify

patterns or themes within qualitative data to answer or explore a research question [25,26]. The analysis was conducted collaboratively by 2 researchers (RT and JY), who followed Braun and Clarke's [25] step-by-step guide to conduct a thematic analysis by familiarizing themselves with the 15 transcripts and coding the data. They then organized the codes based on relatedness, reviewing them, and defining and naming them as subthemes and themes [25]. The thematic analysis acknowledges "the researcher's reflective and thoughtful engagement with their data, and their reflexive and thoughtful engagement with the analytic process is essential" [25]. It recognizes the potential benefits of using multiple coders, such as achieving richer interpretations, however, it does not view this as a requirement [26]. Researchers are discouraged from attempting to provide accounts of "accurate" or "reliable" coding or pursuing consensus among multiple coders [26].

Given the research question and acknowledging that participants' experiences and perspectives on safety may vary in this study, the analysis was used to reflect the range of experiences of participants and highlight how these might differ, rather than attempting to merge these experiences into a single, unified interpretation [25]. Once the results of the study were ready, they were shared with all 15 participants to review and ensure that they were representative of their experiences. Three participants responded and said that they agreed with the results and did not offer any additional insights or suggest any alterations.

Researcher Reflexivity

A critical realist epistemology was adopted for this study, where the researchers aimed to explore participants' subjective experiences, acknowledge them as "real," and recognize the researchers' inability to fully access that reality [25]. The researchers were aware of their reflexivity [27]; at the time of this study, they all worked on a separate clinical trial that aimed to assess the efficacy and safety of a specific DMHI. The first author (RT) was completing her PhD on the safety of DMHIs. This study was 1 of 4 separate pieces of work for the PhD (other work comprising a systematic review, a methodology paper, and an experts' consensus study). This study was not directly related to or in any way part of the clinical trial that researchers were working on. As a team, the researchers were invested in learning how users/patients experience risks, react to them, and what risks matter to them in order to contribute to the field and improve their approach to safety.

Ethical Approval

Ethical Clearance was provided for this study by the King's College London (reference number LRS/DP-22/23-35403).

Results

Overview

The analysis of the data using deductive thematic analysis led to 3 major themes:

- Nonresponse: A Concern, a Risk, and How Users Mitigate It
- Symptom Deterioration and Its Management
- Concerns Around Data Privacy and How to Mitigate Them

Theme 1: Nonresponse: A Concern, a Risk, and How Users Mitigate It

Assessing the Effectiveness of DMHIs

Under this theme, participants spoke about their concerns regarding the DMHI being ineffective, experiencing ineffectiveness/nonresponse as a risk, and the methods they used to assess whether a DMHI was safe and effective.

Concerns Around Nonresponse

Users of DMHIs were concerned about the potential ineffectiveness of these interventions. Will these interventions be able to help them? Are these interventions evidence-based? One important area of concern is illustrated as follows:

> *Umm, I was concerned with like how helpful it would actually be, being that it is an online thing and like I'm not actually talking to a person you know...I was even actually concerned when I started the first session, whether the program would have lasting effects on actually helping me or supporting my mental health.* [Participant 6, used a DMHI for depression for 2.5 months]

It is likely that participants were doubtful about their interventions' effectiveness because they were struggling and in such emotional pain that they were unsure how a technology with no human could alleviate their pain and improve their mood.

Nonresponse as a Risk

Participants also spoke about the risk of the DMHI being unhelpful and ineffective, and how that at times led to further frustration, deterioration, and self-blame, for example:

> *I found it ineffective, if I'm honest. I think it was the set of six or eight weeks...I remember that the second week I burst into tears. It just felt so pointless. I can't remember what set me off, but it just felt so pointless.* [Participant 9 used a DMHI for an eating disorder for 1.5 months]

Another participant explained how the ineffectiveness of the DMHI led to a deterioration in her symptoms and feelings of isolation and self-blame. She said:

> *So, it's sort of added to that frustration when it was making the situation worse...So now, I felt like it would lead me to do sort of negative coping strategies...I'd be like get angry and irritable with people, or I'd go and overeat....it definitely sort of furthered the thoughts that there was like no one to help me...It made low moments even worse...It almost triggered sort of thoughts of like, oh, something's wrong with me. Why can't the program help me?* [Participant 14 used a DMHI for post-natal depression for 24 months]

Users' Method for Assessing Safety and Effectiveness

Users used 2 main methods to assess whether they thought a DMHI was safe and effective: (1) social proofing, which refers

to the tendency to follow the behavior of others as a guide for one's own actions [28]; and (2) assessing the contribution of experts or a trusted body. Some participants opened up about finding it difficult to assess the safety of a DMHI, and not knowing how to do that.

Participants wanted to know that professionals were involved in the development of the DMHI and that scientific research has been conducted to assess it, saying:

> *I suppose I would want to know how it was developed and in partnership with mental health practitioners...so I think with credibility, I guess things like whether they have worked with the university or with kind of recognized academics and done any kind of scientific research and rather than just user testing.* [Participant 11 used a DMHI for insomnia for 1.5 months]

In this participant's case in particular, knowing that the intervention was evidence-based was very important because she had struggled with insomnia for more than a decade and had tried many things (face-to-face therapy and medication) that were not helpful for her. Users also relied on other users' experiences. They checked reviews, ratings, and social media groups to find out more about the DMHI. The following quote is an example of this:

> *Uh, I have this habit of looking up these things online, so I would look up reviews of the app online. Maybe even check out any details about how helpful it has been. If people have had good experiences, bad experiences, they've had negative experiences, what have they been about and how I could avoid them. I would maybe even like look at Facebook groups.* [Participant 6 used a DMHI for depression for 2.5 months]

Some participants were honest about not knowing how to check the safety of the DMHI, and needing the support of professionals to be able to do so, they said:

> *I mean, unless it was like referred by my GP or kind of, you know, promoted through kind of official channels like the NHS website or something, I don't know how I would even check that an app has all the right checks, and you know safeguarding approvals or whatever.* [Participant 13 used a DMHI for depression and anxiety for 2 months]

Other participants suggested that the DMHI needs to assess users' suitability, saying:

> *I guess before someone's able to access the app kind of going through, I don't know, some sort of risk assessment on like who would find it useful.* [Participant 7 used a DMHI for paranoia for 3 months]

Another participant thought that this could be achieved by the DMHI clearly stating its intended use and target population:

> *I think the app would need to be really explicit about the limitations and sort of say up front like this is not for severe mental health issues or this is for maintenance.* [Participant 14 used a DMHI for post-natal depression for 24 months]

Theme 2: Symptom Deterioration and Its Management

Addressing Symptom Deterioration in DMHIs

Under this theme, participants spoke about experiencing symptom deterioration. Participants also gave their feedback on one of the methods used to support them when experiencing or struggling to cope (referral to other services and crisis support) and their suggestions on how deterioration can be managed.

Deterioration of Symptoms

Participants talked about how using a DMHI and dealing with their mental health struggles at times led to a deterioration in their mental health symptoms because it made them think about things that were upsetting. One participant who was struggling with sleep said:

> *I did find thinking more closely about my trouble with sleep did initially make me more anxious about sleep and made it harder to sleep. So, like the kind of tracking and then realizing that actually that was a really bad night...sometimes makes it harder to sleep the next night, by bringing it to the forefront.* [Participant 11 used a DMHI for insomnia for 1.5 months]

It is important to note that this participant found her intervention effective in helping her manage her insomnia and improve her sleep quality. This aligns with recent findings from an expert consensus study, which concluded that symptom deterioration is not a safety concern of DMHIs but rather a normal part of therapy [11].

In some cases, the inflexibility of the predetermined content in the DMHI and its inability to cater specifically to each user's needs (ie, lack of personalization) meant that the DMHI was unable to relate to users' emotional state and could lead to a deterioration in symptoms, for example:

> *So sometimes the AI (chatbot) like would give me suggestions that didn't really fit my situation. So, I'm like, you know what? Forget it. I'm not even going to do it, and I would feel worse afterwards because I wanted to express it. And then I'm just sitting here typing things, and it's not helping. It wasn't really built to recognize that CBT isn't effective for certain situations.* [Participant 14 used a DMHI for post-natal depression for 24 months]

Managing Deterioration

As deterioration is the most common negative effect of DMHIs [1], the authors asked participants how they think DMHIs could support them through it. Some participants said that normalizing deterioration would be very helpful. Participant 1 explained how that could be done in a hybrid model:

> *I think one way to support...is to have a video call therapy session. With an agent for example to make me understand that these things are normal. So, at that I would be reassured that I'm getting back to*

normal. [Participant 1 used a DMHI for depression for 36 months]

Another participant explained how that could be done in a nonhybrid model:

> *Uh, maybe you know if the app had a mood tracker. The algorithm could check if you're feeling low right after therapy. It might just send them a message "Hey, if you're feeling down, you just check, you just had therapy. This could be normal" and that would be like, uh, fair enough.* [Participant 5 used a DMHI for depression and anxiety for 1 month]

Another participant suggested using regular check-ins to detect deterioration and provide users with support accordingly, saying:

> *Umm, I think the biggest thing is regular check-ins. That's quite an important thing about how they're finding it and what particularly is so difficult, maybe even not slowing it down, but having a bit of flexibility around kind of OK, you found this section of the app pretty difficult.* [Participant 10 used a DMHI for self-harm for 1.5 months]

Signposting to Other Services for Support

Most participants (11/15, 73%) were provided with emergency numbers, within the DMHI, to call in case they felt that they could not cope and needed further support. Some participants shared that they found this support helpful:

> *They gave me the contact numbers of like mind and Samaritans in case I needed urgent help. I'm using an online service. If I did need help, I could contact these services, which is actually really helpful because once or twice when I really felt like I was troubled at night, this did come in handy.* [Participant 6 used a DMHI for depression for 2.5 months]

However, other participants felt that the signposting to crisis support within the DMHI was tokenistic, sharing:

> *It's hard to feel like it's a genuine thing. It feels almost like a boilerplate that they put in every conversation. It doesn't really feel like there's thought going into it like "Ohh I recognize that the program can't help you with this. This would be better for a psychiatrist.* [Participant 14 used a DMHI for post-natal depression for 24 months]

Further analysis of these data highlighted that the participants who had a positive experience with the DMHI (ie, found it helpful) experienced the referral to crisis support information positively. By contrast, participants who found the DMHI unhelpful and were frustrated with it found the crisis support information unhelpful and ingenuine. Thus, users' relationship with the DMHI and their feelings toward it informed how they felt about being referred to other services. Users who found the DMHI helpful were likely to view the information about other services as a helpful bonus, whereas those who found their DMHI unhelpful were likely to doubt its genuine concern for them and thus viewed such referrals as a mere box-ticking exercise.

Theme 3: Concerns Around Data Privacy and How to Mitigate Them

Participant Perspectives on Data Privacy

Under this theme, participants spoke about their concerns around data privacy and their suggestions on how to help ease these concerns.

Concerns Around Confidentiality and Data Privacy

Participants were concerned about their data. Was the DMHI confidential? If not, who is it sharing their data with? Participants talked about their concern that the data might be shared with their health care team without their consent:

> *I had concerns about the information that I was putting in, and the kind of data that might have been collected. It's kind of that worry that what you're writing isn't actually confidential or you know that it could go back to someone else. I think one of the biggest worries for me was that what I was inputting in the app might have been given to my psychologist or somehow, you know, connected with the NHS or something like that.* [Participant 10 used a DMHI for self-harm for 1.5 months]

It is worth noting that, given the sensitivity of participant 10's struggle with self-harm, it is understandable why they were particularly concerned about their data being shared—even with their health care provider.

A User-Friendly Data Policy

Participants expressed their frustration with the vagueness and complexity of how DMHIs present their data protection policy and had suggestions on how that can be improved. Some thought that DMHIs should make key data protection information available in a simpler and more readily accessible format:

> *Maybe just share more information...like make it clear what you do to protect users' data. I don't want to have to go through, you know, all of your Terms and conditions, privacy policies and things like that to find out what it is. I mean, nobody's actually going to do that. I would never actually do it, so it would be helpful if it was just clearly mentioned somewhere.* [Participant 6 used a DMHI for depression for 2.5 months]

Another participant said that all they wanted from a DMHI is to be honest about what data they are storing and why:

> *They could kind of like emphasize maybe that your data isn't stored under an identifiable name that leads to you. Or say we are storing the data, but it is confidential and we're doing it to help more people. That's all...just being open about it.* [Participant 8 used a DMHI for an eating disorder for 3 months]

It is worth noting that the data from this study did not show any association between duration of use and participants' experiences. Participants had mixed responses to the DMHI when they used it for a short period (1.5 months, eg, participants 10 and 11), and those who used the DMHI for a long period (24 months; eg, participant 14) did not necessarily have a good

experience. However, Table 1 does show that the DMHIs that were targeting low mood–related areas were used for the longest period, such as depression (on average 19.25 months), self-harm (on average 12.75 months), and postnatal depression (on average 24 months). It is also noted that self-administered interventions were used for longer periods compared with hybrid interventions (on average 9.33 months vs 1.66 months). This is expected given that self-administered interventions require fewer resources and clinician time.

Discussion

Advancing Knowledge on DMHI Safety

This research was successful in exploring and understanding users' experiences, views, concerns, and suggestions regarding the safety of DMHIs. Until now, such findings have been absent from the literature. These findings contribute to advancing the field of digital mental health safety by providing valuable evidence of the viewpoints and experiences of its target population.

Principal Findings

Overview

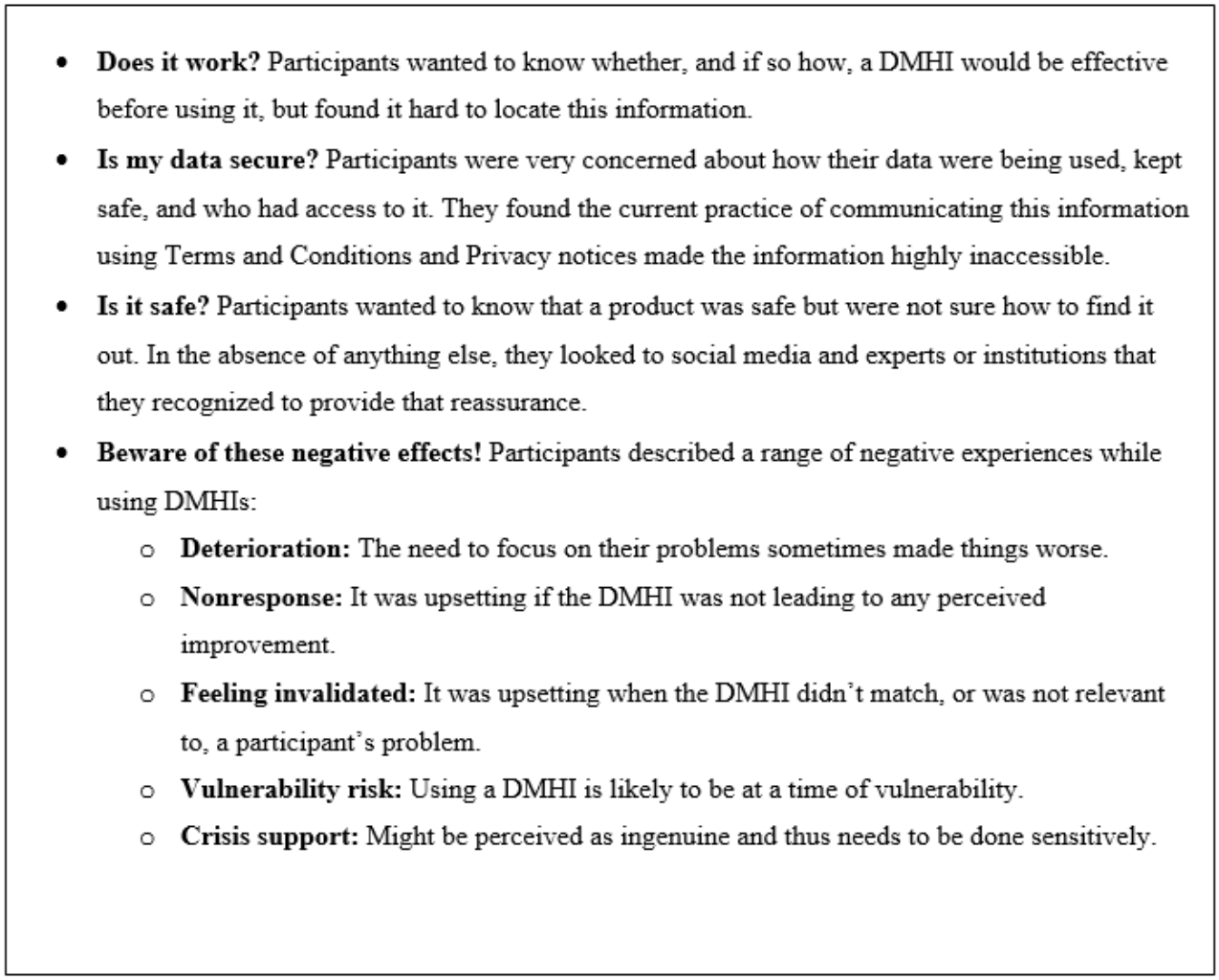
The main findings of this study are presented in Table 2 (also see Figure 1) using user-friendly language and in the form of recommendations.

Table 2. User-informed recommendations to improve DMHI[a] safety.

Recommendations	Description
1. Provide 'easy access' versions of key information	Ensure that DMHI product manufacture and approval include a requirement to provide readily accessible, easy-to-read lay summaries of key information. At a minimum, these should cover (1) evidence of effectiveness; (2) data usage, security measures, and access (ie, who can access the data); and (3) potential negative effects.
2. Use 'Approved by...' badges	Introduce a sectorwide, widely recognized, branded badge to provide top-level reassurance of the quality and safety of any DMHIs bearing that badge.
3. Acknowledge the lack of personalization	DMHIs should flag to users their inability to be fully personalized and adaptable to an individual user's needs to mitigate feelings of invalidation and disappointment.
4. Anticipate and support deterioration	Before using a product, users should be alerted to possible mood or symptom deterioration, given normalizing information, and signposted to relevant support to help mitigate these effects should they occur.
5. Provide real-time feedback	DMHIs should internally track users' progress and provide feedback on whether they are benefiting as expected. Where there is no individual benefit, despite the appropriate use of the product, users should be automatically advised to seek alternative support.
6. Provide genuine crisis support	Content should include an acknowledgment of the DMHIs' limitations and a summary of each crisis service's support. It is recommended to consult the target population on wording to ensure genuine concern for users is communicated.
7. Responsibly manage access	DMHIs should incorporate an assessment of suitability focusing on risk levels and the appropriateness of the intervention for users' specific mental health conditions and severity. This could be done with a simple set of initial built-in questions which output a recommendation to use, or not use, the product based on the user's response.

[a]DMHI: digital mental health intervention.

Figure 1. Key findings (in lay language). DMHI: digital mental health intervention.



Users' Concerns Around the Safety of DMHIs

In this study, users expressed 2 primary concerns regarding the safety of DMHIs: (1) whether a product would be effective and (2) whether their data would be secure and confidential. Evidence around the effectiveness of DMHIs is usually disseminated in academic peer-reviewed articles. This poses an accessibility challenge for the typical users, which is only slightly tempered by recent initiatives toward making open-access publications the norm in academia. It has been documented that uncertainty around the effectiveness of a DMHI is a key barrier to its use [15]. To address users' concerns about effectiveness, it is crucial to translate scientific findings into lay language and publish them in user-friendly formats to improve accessibility (recommendation 1). This dovetails well with the increasing requirements placed upon academics to evidence the impact of their work.

Regarding users' second major concern, data safety and confidentiality, participants in this study suggested that DMHIs provide their users with clear and concise information on how their data are used, stored, and who has access to it. Recognizing that users often overlook traditional lengthy privacy policies and terms and conditions [29], DMHIs could supplement these by providing users with a layperson's summary of how their data are being used, kept safe, and who has access to them (recommendation 1). Clear and transparent communication about data and privacy would help build trust, a key component of the therapeutic relationship [30]. The evidence to date suggests that the digital therapeutic alliance is both relevant and important in DMHIs [19,20].

Assessing a DMHI's Safety From Users' Perspective

It is important to understand how users assess the safety of DMHIs, as this will inform professionals about where users look for safety information and so where and how best to provide it. Participants in this study described methods for assessing the safety of a DMHI that reflected social proof, a concept first attributed to Robert Cialdini [31,32]. Social proof refers to situations where people use opinions and information from others similar to themselves to influence personal choices, decisions, and behaviors, especially if uncertain [31,32]. In the present context, this involved reading online product reviews on websites and mobile app stores. Other participants expressed that they did not know how to assess the safety of a DMHI. In a different qualitative study, medical students shared a similar experience, expressing difficulty in identifying which DMHIs were evidence-based [15]. These students also criticized the lack of guidance available for users on how to find evidence-based DMHIs [15].

There are already regulatory bodies equipped and responsible for assessing the efficacy and safety of DMHIs and making recommendations for use. In the United Kingdom, this includes NICE (The National Institute for Health and Care Excellence) and the MHRA (Medicines and Healthcare Products Regulatory Agency). The NHS (National Health Service) previously had a health app store called the "NHS Health Apps Library," but it was decommissioned in 2021 due to the increasing complexity of maintaining the library and ensuring the safety and effectiveness of the listed apps. An alternative to this has emerged through the Health App Library provided by the Organization for the Review of Care and Health Apps (ORCHA), in collaboration with health providers such as NHS Trusts. This library offers a list of mobile apps that have been reviewed by ORCHA for effectiveness and safety. One effective way to demarcate a product's safety and efficacy status would be for regulators to introduce a branded, recognizable stamp or badge to be displayed by products achieving prespecified minimum safety and efficacy requirements (recommendation 2). What those requirements should be, however, would entail significant additional research to achieve a consensus across industry, regulatory bodies, academia, developers, and users. Nevertheless, this approach would capitalize on users' existing tendency to seek safety information via the product's mobile app page or website and is therefore likely to ultimately be an effective means of disseminating key information on which end users can base their decisions. A notable advancement in this direction is Google's revision of their app store's health policy, which mandates that starting from May 31, 2024, all health apps posted on their app store must prove compliance with relevant laws and regulations (privacy policy, ethics approval, and certification when required) [33].

Risks Experienced by DMHIs' Users and Their Suggestions on How to Mitigate Them

Participants in this study spoke about 3 risks they had experienced as a result of using a DMHI: feeling invalidated by the DMHI, deterioration in their symptoms, and nonresponse. Users noted that, unlike a human therapist, the DMHIs' inability to be fully responsive, personalized, and adaptable to each user's needs left them feeling invalidated and unheard. Such experiences undermine the therapeutic relationship. There is evidence that personalization in a DMHI fosters therapeutic alliance [20], and thus the lack of it is likely to undermine this alliance. The weaker the digital therapeutic relationship, the more this is likely to undermine the effectiveness of the DMHI [30]. In a qualitative study involving Australian psychologists and their experiences with DMHIs, the psychologists expressed the view that DMHIs are inferior to face-to-face therapy due to their limited capacity for personalization [14]. In another qualitative study, medical students made the same comparison between DMHIs' and health professionals' ability to provide personalized therapy [15]. This is a limitation of current technologies, which might change with the future advances of artificial intelligence. However, for now, it is important for DMHIs to acknowledge and communicate this limitation to users to mitigate feelings of invalidation (recommendation 3).

Participants suggested that DMHIs can support users experiencing deterioration by informing them about the possibility, normalizing it, and providing pathways to relevant support when it occurs (recommendation 4). These suggestions align with those made by digital mental health professionals in a recent consensus statement [11]. Other studies have suggested implementing an automated process within DMHIs to monitor and flag when participants' symptoms deteriorate beyond a predefined threshold [5]. Although that threshold would be subject to individual conditions and clinical opinion within any specific context, a useful starting point would be to adopt the clinical "rule of thumb" that considers a 20% change in symptoms as a meaningful variation [34].

"Nonresponse" is a documented potential side effect of DMHIs [1,12]. It was interesting to see how nonresponse from users' perspective was almost a compound negative effect that led to a cascade of unwanted effects including feelings of frustration, hopelessness, deterioration, isolation, and self-blame. One way to address this would be for DMHIs to track users' clinical outcomes, identify those experiencing nonresponse, and provide them with targeted pathways to further support as a way of mitigating these possible adverse consequences (recommendation 5).

The Risk Mitigation Methods Experienced by DMHIs' Users

When discussing how DMHIs mitigate risks and safeguard users, users highlighted the importance of signposting to other sources of support. The majority of users (11/15) were provided with details of emergency numbers and other mental health services; however, not all users found these helpful. Some users felt that this was a checkbox exercise that DMHIs needed to complete and that the crisis support provided was ingenuine. Users of DMHIs may feel this way because of their experiences with mental health services, and because most mental health services/interventions tend to provide crisis support information. Signposting to a different service needs to be done delicately to ensure that the user feels cared for. For that, DMHIs need to be careful about how they present crisis support details. Including an acknowledgment of the DMHI's limitations, a concern for the user, and a description of the support that each crisis service provides might help make users feel that the intention to provide support is more genuine (recommendation 6). Soliciting user input on how such information is phrased and presented would also be of benefit (recommendation 6).

Only 1 user among our sample of 15 was informed of the potential side effects of the DMHI that they were using. A recent systematic review on the safety of DMHIs found that only one-third of the interventions informed their users of their adverse events or possible side effects [1]. The even lower level of side effect awareness in our study might be due to the commonplace practice of embedding side effect information within inaccessible or often unread documentation (eg, terms and conditions or instructions for use) [6]. Indeed, as reported above, our participants told us that they found these documents particularly impenetrable, suggesting some may have missed out on important side effect information. To ensure user safety, it is important to improve the visibility and accessibility of side effect information by adopting new methods of communication (recommendation 1). This might include a digital equivalent of

listing possible side effects on medication labels. Existing regulations already require a digital product label to be displayed within the product itself and this label includes "Cautions" and "Warnings." It would be a relatively simple matter to add a section "Possible Side Effects" as a further requirement.

However, simply adding information is unlikely, on its own, to meet users' needs as identified by our study. In addition (and as already discussed earlier), serious consideration should be given to adopting the practice of requiring "easy read" or "lay summary" versions of key information which is provided alongside full and formal versions. This is now standard practice in domains such as academia, governmental, and other public sector organizations. Our study suggests that, as a minimum, this should apply to information on data security and side effects.

Finally, to minimize this risk, our sample of participants/users recommended that DMHIs should evaluate each user's suitability, with a focus on assessing risk levels and determining the appropriateness of the intervention for their specific mental health condition and severity. These assessments can be included as a standard procedure before onboarding a user onto a DMHI, similar to how patients are screened before they receive face-to-face therapy (recommendation 7).

Limitations

There are a few limitations to this study. The participants in this study were mostly female (12/15, 80%). The DMHIs used by participants were mostly self-administered (12/15, 80%), and thus results might be biased by their experiences. Additionally, recruiting for a study to explore users' perspectives on the safety of DMHIs might have attracted individuals who have experienced such issues. It is important to be aware of how the sample of participants in this study could have shaped the results. This is expected in qualitative studies, which aim to explore and understand the experiences and opinions of a sample of the population [35].

Conclusions

The results of this study led to 7 user-informed recommendations on how the safety of DMHIs can be improved. These recommendations arose from users' experiences and suggestions. The key findings (Figure 1) and recommendations of this paper could improve the safety of DMHIs, enhance users' experience, address some of their concerns, and foster a more trusting therapeutic relationship between the user and the DMHI.

Acknowledgments

We are grateful to all our participants for taking the time to participate in our study and share with us their valuable experiences and opinions. We also would like to express our gratitude to the National Institute for Health and Care Research (NIHR) Biomedical Research Centre hosted at South London and Maudsley NHS Foundation Trust in partnership with King's College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care, the ESRC, or King's College London.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Topic guide.
[DOCX File , 24 KB - humanfactors_v12i1e62974_app1.docx]

References

1. Taher R, Hsu C, Hampshire C, Fialho C, Heaysman C, Stahl D, et al. The safety of digital mental health interventions: systematic review and recommendations. JMIR Ment Health 2023 Oct 09;10:e47433 [FREE Full text] [doi: 10.2196/47433] [Medline: 37812471]
2. Brown P, Prest B, Miles P, Rossi V. The development of National Safety and Quality Digital Mental Health Standards. Australas Psychiatry 2022 Apr 27;30(2):154-157. [doi: 10.1177/10398562211042361] [Medline: 34569321]
3. van Orden ML, Kraaijeveld JC, Spijker AT, Silven AV, Bonten TN, Chavannes NH, et al. Preliminary effects of a digital mental health intervention for depression and anxiety. Clinical eHealth 2022 Dec;5:44-51. [doi: 10.1016/j.ceh.2022.06.002]
4. Kambeitz-Ilankovic L, Rzayeva U, Völkel L, Wenzel J, Weiske J, Jessen F, et al. A systematic review of digital and face-to-face cognitive behavioral therapy for depression. NPJ Digit Med 2022 Sep 15;5(1):144-144 [FREE Full text] [doi: 10.1038/s41746-022-00677-8] [Medline: 36109583]
5. Gómez Bergin AD, Valentine AZ, Rennick-Egglestone S, Slade M, Hollis C, Hall CL. Identifying and Categorizing Adverse Events in Trials of Digital Mental Health Interventions: Narrative Scoping Review of Trials in the International Standard Randomized Controlled Trial Number Registry. JMIR Ment Health 2023 Feb 22;10:e42501. [doi: 10.2196/42501]
6. Martinez-Martin N, Kreitmair K. Ethical issues for direct-to-consumer digital psychotherapy apps: addressing accountability, data protection, and consent. JMIR Ment Health 2018 Apr 23;5(2):e32 [FREE Full text] [doi: 10.2196/mental.9423] [Medline: 29685865]

7. Berk M, Parker G. The elephant on the couch: side-effects of psychotherapy. Aust N Z J Psychiatry 2009 Sep 01;43(9):787-794. [doi: 10.1080/00048670903107559] [Medline: 19670051]
8. Batterham PJ, Christensen H, Calear AL, Werner-Seidler A, Kazan D. Rates and predictors of deterioration in a trial of internet-delivered cognitive behavioral therapy for reducing suicidal thoughts. Archives of Suicide Research 2020 Nov 20;26(2):937-947. [doi: 10.1080/13811118.2020.1848671] [Medline: 33215554]
9. Linden M. How to define, find and classify side effects in psychotherapy: from unwanted events to adverse treatment reactions. Clin Psychol Psychother 2013 Jan 18;20(4):286-296. [doi: 10.1002/cpp.1765] [Medline: 22253218]
10. Terry NP, Gunter TD. Regulating mobile mental health apps. Behav Sci Law 2018 Mar 16;36(2):136-144 [FREE Full text] [doi: 10.1002/bsl.2339] [Medline: 29659069]
11. Taher R, Bhanushali P, Allan S, Alvarez-Jimenez M, Bolton H, Dennison L, et al. Bridging the gap from medical to psychological safety assessment: consensus study in a digital mental health context. BJPsych Open 2024 Jun 03;10(4):e126-e126 [FREE Full text] [doi: 10.1192/bjo.2024.713] [Medline: 38828683]
12. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of internet interventions. Internet Interventions 2014 Mar;1(1):12-19. [doi: 10.1016/j.invent.2014.02.001]
13. Bradstreet S, Allan S, Gumley A. Adverse event monitoring in mHealth for psychosis interventions provides an important opportunity for learning. J Ment Health 2019 Oct 26;28(5):461-466 [FREE Full text] [doi: 10.1080/09638237.2019.1630727] [Medline: 31240970]
14. Scott S, Knott V, Finlay-Jones AL, Mancini VO. Australian psychologists experiences with digital mental health: a qualitative investigation. J Technol Behav Sci 2022 Aug 16;8(4):1-11 [FREE Full text] [doi: 10.1007/s41347-022-00271-5] [Medline: 35991293]
15. Dederichs M, Weber J, Pischke CR, Angerer P, Apolinário-Hagen J. Exploring medical students' views on digital mental health interventions: a qualitative study. Internet Interv 2021 Sep;25:100398 [FREE Full text] [doi: 10.1016/j.invent.2021.100398] [Medline: 34026567]
16. Allan S, Beedie S, McLeod HJ, Farhall J, Gleeson J, Bradstreet S, et al. Using EMPOWER in daily life: a qualitative investigation of implementation experiences. BMC Psychiatry 2023 Aug 17;23(1):597-597 [FREE Full text] [doi: 10.1186/s12888-023-05096-x] [Medline: 37592231]
17. Arnold C, Williams A, Thomas N. Engaging with a web-based psychosocial intervention for psychosis: qualitative study of user experiences. JMIR Ment Health 2020 Jun 19;7(6):e16730 [FREE Full text] [doi: 10.2196/16730] [Medline: 32558659]
18. Berry N, Lobban F, Bucci S. A qualitative exploration of service user views about using digital health interventions for self-management in severe mental health problems. BMC Psychiatry 2019 Jan 21;19(1):35-13 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
19. Tong F, Lederman R, D'Alfonso S, Berry K, Bucci S. Conceptualizing the digital therapeutic alliance in the context of fully automated mental health apps: a thematic analysis. Clin Psychol Psychother 2023 Apr 11;30(5):998-1012. [doi: 10.1002/cpp.2851] [Medline: 37042076]
20. Tong F, Lederman R, D'Alfonso S, Berry K, Bucci S. Digital therapeutic alliance with fully automated mental health smartphone apps: a narrative review. Front Psychiatry 2022 Jun 22;13:819623-819612 [FREE Full text] [doi: 10.3389/fpsyt.2022.819623] [Medline: 35815030]
21. Braun V, Clarke V. Successful Qualitative Research A Practical Guide for Beginners. London, UK: SAGE Publications Ltd; Mar 01, 2013.
22. Moltu C, Stefansen J, Svisdahl M, Veseth M. Negotiating the coresearcher mandate - service users' experiences of doing collaborative research on mental health. Disabil Rehabil 2012;34(19):1608-1616. [doi: 10.3109/09638288.2012.656792] [Medline: 22489612]
23. Braun V, Clarke V. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. Qualitative Research in Sport, Exercise and Health 2019 Dec 26;13(2):201-216. [doi: 10.1080/2159676x.2019.1704846]
24. NVivo (version12). QSR International Pty Ltd. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2025-01-31]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. Byrne D. A worked example of Braun and Clarke’s approach to reflexive thematic analysis. Qual Quant 2021 Jun 26;56(3):1391-1412. [doi: 10.1007/s11135-021-01182-y]
27. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qualitative Research in Sport, Exercise and Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
28. Scott W, Barden J. Social Proof. London, UK: Routledge; Apr 22, 2022:36-64.
29. Wykes T, Lipshitz J, Schueller SM. Towards the design of ethical standards related to digital mental health and all its applications. Curr Treat Options Psych 2019 Jul 5;6(3):232-242. [doi: 10.1007/s40501-019-00180-0]

30. Boucher EM, Harake NR, Ward HE, Stoeckl SE, Vargas J, Minkel J, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices 2021 Dec 31;18(sup1):37-49 [FREE Full text] [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]
31. Cialdini RB. Influence, New and Expanded: The Psychology of Persuasion. New York, NY: HarperCollins; Oct 27, 2013.
32. Cialdini R. Influence: Science and Practice. London, UK: Pearson Publishing; 2009.
33. Google LLC. Health content and services. Google LLC. URL: https://support.google.com/googleplay/android-developer/answer/13996823?sjid=11315650021394300636-EU [accessed 2025-01-31]
34. Hobbs C, Lewis G, Dowrick C, Kounali D, Peters TJ, Lewis G. Comparison between self-administered depression questionnaires and patients' own views of changes in their mood: a prospective cohort study in primary care. Psychol Med 2020 Jan 20;51(5):853-860. [doi: 10.1017/s0033291719003878]
35. Leung L. Validity, reliability, and generalizability in qualitative research. J Family Med Prim Care 2015;4(3):324-327 [FREE Full text] [doi: 10.4103/2249-4863.161306] [Medline: 26288766]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DMHI: digital mental health intervention
MHRA: Medicines and Healthcare Products Regulatory Agency
NHS: National Health Service
NICE: The National Institute for Health and Care Excellence
ORCHA: Organization for the Review of Care and Health Apps
PIS: participant information sheet

Edited by A Kushniruk; submitted 06.06.24; peer-reviewed by I Wilson, L McCann; comments to author 14.10.24; revised version received 07.11.24; accepted 24.11.24; published 07.02.25.

Please cite as:
Taher R, Stahl D, Shergill S, Yiend J
The Safety of Digital Mental Health Interventions: Findings and Recommendations From a Qualitative Study Exploring Users' Experiences, Concerns, and Suggestions
JMIR Hum Factors 2025;12:e62974
URL: https://humanfactors.jmir.org/2025/1/e62974
doi:10.2196/62974
PMID:

Original Paper

Provider Perspectives on the Use of Mental Health Apps, and the BritePath App in Particular, With Adolescents at Risk for Suicidal Behavior: Qualitative Study

Frances Lynch[1*], PhD, MSPH; Julie Cavese[1,2*], PsyD, MA; Lucy Fulton[1*], AA; Nancy Vuckovic[3*], PhD; David Brent[4*], MD, MS

[1]Kaiser Permanente Center for Health Research, Portland, OR, United States

[2]Cavese Counseling, Portland, OR, United States

[3]Vuckovic Consulting, Vuckovic Consulting, Portland, OR, United States

[4]Department of Psychiatry, University of Pittsburgh Medical Center, Pittsburgh, PA, United States

[*]all authors contributed equally

Corresponding Author:
Frances Lynch, PhD, MSPH
Kaiser Permanente Center for Health Research
3800 N Interstate Ave
Portland, OR, 97227
United States
Phone: 1 503 335 2400
Fax: 1 503 335 2428
Email: frances.lynch@kpchr.org

Abstract

Background: Many youth with significant mental health concerns face limited access to mental health services. Digital programs, such as mobile apps designed to address mental health issues, have the potential to expand access to strategies for managing these conditions. However, few mental health apps are specifically designed for youth experiencing severe concerns, such as suicidal ideation. BritePath is a new app developed to enhance communication and interaction between providers and youth at risk for suicidal behavior.

Objective: This study aims to explore health care providers' opinions and concerns regarding the use of mental health apps for youth at significant risk of suicidal behavior.

Methods: We conducted individual semistructured interviews with 17 providers across 7 states. Interviews were conducted via video, recorded, and transcribed. Codes were developed using a team-based approach, with discrepancies resolved through team discussions.

Results: Most providers were aware of mental health apps in general and expressed interest in trying the BritePath app with patients experiencing depression, suicidality, or both. Analyses identified 4 key themes related to mental health apps: (1) almost all providers viewed mental health apps as an adjunct to, rather than a replacement for, psychotherapy visits; (2) most providers were concerned about the cost of apps and youth access to them; (3) providers noted the challenge of maintaining patient engagement with apps over time; and (4) providers were concerned about patient privacy, in terms of both data shared with app developers and data privacy within families. Analyses of providers' opinions specifically about the BritePath app identified 4 additional themes: (1) providers believed that access to safety plans within BritePath could be beneficial for youth at risk for suicidal behavior; (2) providers reported that BritePath's interactive features could enhance communication between providers and youth; (3) providers appreciated BritePath's flexibility and the ability for both youth and providers to tailor its content to individual needs; and (4) providers expressed concerns about integrating BritePath into clinical workflows within health systems.

Conclusions: The use of mental health apps is expanding, yet there is limited understanding of how to effectively integrate these tools into mental health treatment. Providers are increasingly referring patients to mental health apps, and most expressed interest in trying the BritePath app for patients with depression, suicidality, or both. However, providers also identified several concerns, particularly regarding privacy and safety.

(JMIR Hum Factors 2025;12:e64867) doi:10.2196/64867

KEYWORDS

depression; adolescent; suicidality; safety plan; mental health; apps; suicide

Introduction

Background

Adolescent suicidal behavior, suicide ideation, and depression are major public health problems that have increased significantly over the past 20 years [1-3] and have been exacerbated by the COVID-19 pandemic and its sequelae [4-8]. For example, between 2019 and 2021, the number of female high school students reporting that they had seriously considered attempting suicide increased by 6% [3]. These increases in suicidal ideation, behavior, and depression have disproportionately affected youth from historically minoritized racial and ethnic groups [2-4]. Once youth are identified as having a significant risk of suicidal behavior, they are typically treated in emergency departments or urgent care settings or hospitalized for stabilization before beginning outpatient treatment [9]. However, many youth remain at high risk of suicidal behavior even after acute treatment or stabilization in intensive settings. Managing youth with high levels of suicidal risk requires close communication between patients and clinical providers, such as psychiatrists or mental health therapists (referred to collectively as providers hereafter). Evidence-based care for these youth includes regular monitoring of depression symptoms, suicidal ideation, and suicidal behavior [9-11]. However, most health systems fall short of maintaining this level of close communication due to barriers such as lack of time and resources, as well as difficulties in staying in touch with at-risk youth after they leave intensive settings or acute treatment.

To address these concerns, researchers are developing new mobile apps aimed at improving both the efficiency of communication with youth at risk for suicidal behavior and shared decision-making between providers and adolescent patients. These tools are increasingly promoted to address mental health concerns in general and depression in particular [12-16]. Empirically based apps for depression range from cognitive behavioral self-help programs [15,16] to screening [16,17] or mood-tracking apps [18]. Recent research indicates that these tools are generally acceptable to providers [13,17]. However, there are several limitations to their use in clinical practice. Much of the evidence comes from surveys asking providers about the acceptability of apps [19] or studies that have not focused on specific concerns about particular apps [13]. Few studies have examined the use of apps within the context of mental health treatment [15]; instead, most have focused on patients' use of apps independent of providers [13]. To date, no prior research has examined the acceptability of apps designed to support shared decision-making between providers and youth at high risk for suicidal behavior. While some experts have raised potential ethical and safety concerns about using digital interventions with high-risk populations [19,20], improving connections to treatment after nonfatal suicidal behavior has the potential to enhance the treatment trajectory and long-term outcomes for these youth.

Study Goal

The aims of this study were to characterize providers' (1) opinions on the barriers and benefits of using apps in mental health care in general, (2) perspectives on the use of apps with adolescents with depression or suicidal thoughts and young adult patients, (3) interest in and willingness to use a recently developed app (BritePath) intended for use with adolescents with suicidal thoughts or young adult patients, and (4) barriers to implementing the BritePath app in routine clinical practice.

Methods

Implementation of BritePath: Benefits and Barriers

This qualitative study was conducted as part of the Center for Enhancing Treatment and Utilization for Depression and Emergent Suicidality (ETUDES Center), an ALACRITY Center funded by the National Institute of Mental Health (NIMH P50MH115838), aimed at helping providers better support youth experiencing severe depression, suicidality, or both. This study explores issues related to the use of mental health apps in general with this population and examines potential benefits and barriers to implementing an already developed app, BritePath, in health systems that have not participated in the research studies developing BritePath and are not affiliated with the ETUDES Center. To date, all information regarding the acceptability of the BritePath app has been gathered within the institutions that developed it.

The BritePath App

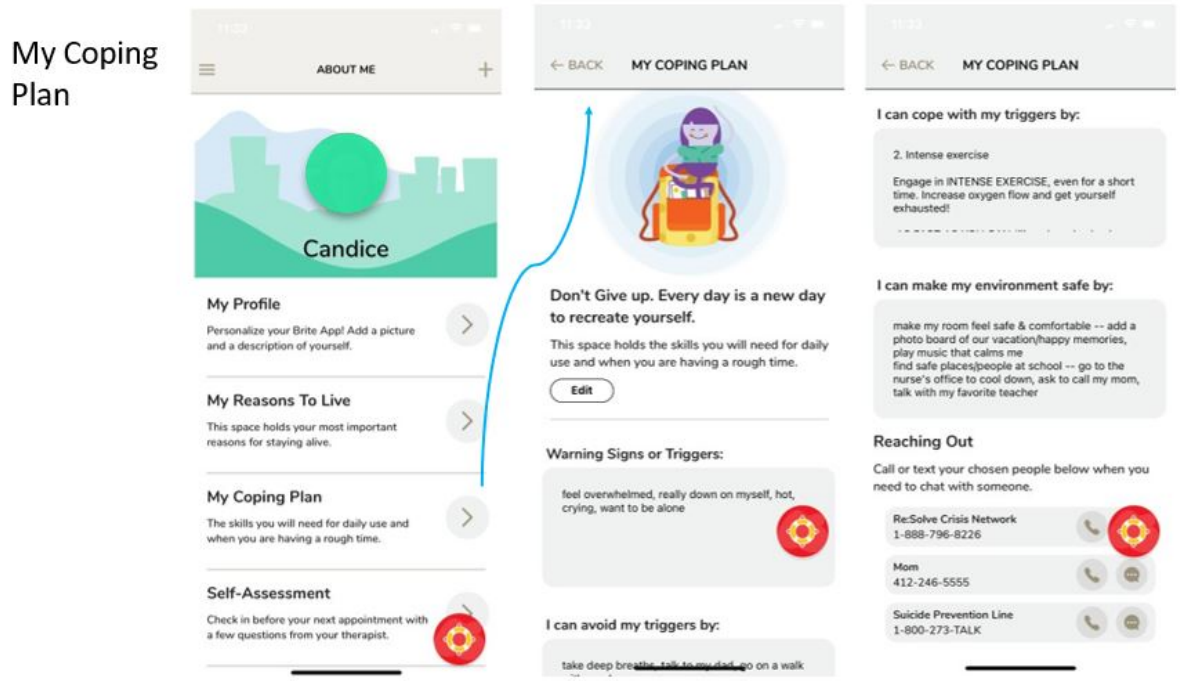
The BritePath app was developed to improve communication between providers and youth with depression or suicidal thoughts. It consists of 3 integrated components: (1) *Guide2BRITE*, an electronic guide for mental health providers that includes step-by-step instructions for onboarding a patient's safety plan into the app, as well as guidance on discussing key treatment components such as emotion regulation and distress tolerance skills; (2) the *BRITE app*, a personalized and interactive safety plan and self-monitoring tool for youth; and (3) the clinician dashboard, *BRITEBoard*, which allows providers to track youth's app use, distress levels, and treatment progress while facilitating communication and collaboration among mental health and primary care providers.

BritePath promotes self-monitoring and self-management through personalized strategies to avoid or cope with triggers for suicidal urges [21]. It is based on BRITE, a patient-facing safety planning app developed by researchers in psychiatry and psychology at the University of Pittsburgh and the University of Texas Southwestern Medical Center [21]. To our knowledge, BritePath is the first smartphone-based safety planning app to be tested in clinical trials involving adolescents with severe depression or suicidal thoughts. BritePath provides providers

with tools to onboard their adolescent and young adult patients to the app. This includes assisting youth in developing and personalizing a safety plan, incorporating distress tolerance and emotion regulation strategies, and monitoring symptomatic progress for at-risk youth in their care. Figure 1 provides a screenshot from the BritePath app for a mock patient.

BritePath was first tested for clinical effectiveness in a pilot randomized trial that evaluated BritePath in combination with a related intervention. The study found that the combined intervention reduced the rate of suicide attempts by 50% in the 6 months following hospital discharge [21]. This pilot study also suggested that more frequent use of BritePath increased reasons for living. However, the small sample size did not allow for comprehensive testing of BritePath alone. A more recent, larger randomized trial found that BritePath is as effective as usual care for youth hospitalized for suicidality. Additionally, youth who received BritePath were less likely to have a subsequent suicide attempt and had a longer time until a repeat attempt [22].

Figure 1. Example of the safety plan in the BritePath app.



Sampling and Recruitment

We identified a convenience sample of mental health and primary care providers who have cared for patients with mental health concerns. Providers were recruited through health systems participating in the Mental Health Research Network [23] and via snowball sampling [24,25], in which interviewed providers recommended additional potential participants. No specific level of digital experience or skills was required for participation in the study.

Ethical Approval

All study procedures were approved by the Institutional Review Board of the Kaiser Permanente Center for Health Research (KPCHR) at Kaiser Permanente Northwest (approval ID: MOD20050381-003), where the study was conducted. Interviews were conducted via KPCHR Teams, and digital recordings and transcripts were securely stored within the primary organization's firewall-protected file service.

Data Collection

Semistructured interviews were conducted remotely via Microsoft Teams. The interview guide (Multimedia Appendix 1) was developed by our medical anthropologist expert (NV) in coordination with the principal investigator (FL). Additionally, a qualitative investigator from the larger BritePath team reviewed the guide for content. The interviewer (FL) used a semistructured interview guide (Multimedia Appendix 1) to ensure consistency across interviews. The framework that guided the design of our qualitative interview guide focused on usability and concept testing. We examined whether providers from different backgrounds and training would consider using BritePath. The included questions aligned with several frameworks for adopting new technology, such as the Diffusion of Innovation theory [26,27]. The interviews assessed providers' opinions on using apps with patients and their impressions of the benefits and barriers of mental health apps in general. Participants were then shown a series of visuals (screenshots) depicting what a user would see while interacting with the BritePath app. These prompts ensured that participants had a clear and consistent understanding of the app's features and user interface. During this portion of the interview, participants were asked about their initial impressions of the BritePath app, as well as the barriers and potential benefits of using it in their practice. They were also given the opportunity to suggest features they would like to see in future app developments. Each interview lasted 30-40 minutes and was audio recorded and transcribed by an independent transcriptionist. Transcripts were entered into Atlas.ti [28], a qualitative analysis software program

used for coding and managing data. Atlas.ti was then used to generate reports for the analysis of interview transcripts.

Analysis

Thematic analysis, a method for analyzing patterns of meaning in a data set, was used to interpret the data by identifying themes and their interconnections [29-32]. Following established thematic analysis procedures [30], members of the research team (JC, LF, and NV) first familiarized themselves with the data and, together with FL, developed an initial code list consisting of a priori codes reflecting the interview questions. JC and LF coded the first interview using this initial code list and identified additional codes while reviewing the transcripts. The full team met to review coding consistency between reviewers and to evaluate the need for new, emergent codes. Modifications to the code list were made based on discussion and consensus. This process was repeated, with 2 team members coding 2 additional transcripts each. The full team met to discuss differences between coders for these additional transcripts and found minimal discrepancies in the codes applied. Upon confirming sufficient consistency in code usage and identifying no additional emerging codes, a master code list was finalized and used to code the remaining transcripts. After coding all transcripts, the full study team individually reviewed reports of coded interview segments to identify themes. The final themes were then derived through discussion and consensus. This formal, team-based analysis procedure was used to enhance the credibility and trustworthiness of the findings.

Results

Providers' Perspectives on Mental Health Apps and the BritePath App for Youth at Risk of Suicidal Behavior

We conducted semistructured interviews with a convenience sample of 17 providers from 7 states in the United States (Table 1). Our goal was to understand providers' opinions on the use of apps for youth with mental health concerns in general and the use of the BritePath app for youth at risk of suicidal behavior. First, we asked providers about their experiences using mental health apps with patients. Next, we gathered their opinions on a specific app, BritePath, designed to improve patient engagement and communication with providers for youth at risk of suicidal behavior. The interview guide is available in Multimedia Appendix 1.

Table 1. Provider demographics (N=17).

Characteristic	Values, n (%)
Age (years) group	
25-34	4 (23)
35-44	6 (35)
45 and over	7 (41)
Sex at birth	
Female	15 (88)
Male	2 (22)
Provider type	
Physician	2 (22)
Therapist/social worker	8 (47)
Psychologist	7 (41)
Type of setting	
Large health system	10 (22)
Smaller group practice	4 (47)
Individual practice	3 (41)

Providers' Opinions About the Use of Apps for Mental Health Concerns in General

Overview

All providers reported familiarity with mental health apps, and most stated that they had referred at least one patient to a mental health–focused app—for example, to help patients learn a skill such as relaxation. However, none of the providers reported regularly using any mental health app during treatment visits or in an interactive manner with patients. Each theme is discussed in more detail below.

Theme 1: Mental Health Apps Have the Potential to Be Valuable as an Adjunct to Therapy Visits

Providers reported that they felt mental health apps could add value to the therapeutic relationship as an adjunct or complement to face-to-face or video-based psychotherapy sessions. However, most did not believe that mental health apps could serve as a substitute for psychotherapy visits. Several providers noted that these apps can offer access to exercises for learning new skills or encourage the use of tools discussed in therapy, making them a useful complement to treatment sessions.

> *I think they can be really great adjuncts if you have the time to utilize those as a clinician. I am always interested for my patients of course for them doing things in the many hours outside of therapy, so I see you from 1:00 to 2:00 and then there are all these other hours that in theory you're doing the things that we were practicing in that one hour. If there are apps that can promote, engage, and promote the use of those strategies, whether it's thought records or sleep diaries or mood meters, things like that. And then they can bring that back in and we can look at it together and that facilitates the conversation, I think that would be great.* [P5]

> *And this way as a clinician I am teaching a skill, but I want it to be reinforced by this tool that mainly my client really loves but eventually is a way to train the brain to think a new and different way.* [P4]

Providers also noted that this approach to learning skills may be particularly useful for teenagers, who frequently engage with their phones.

> *I think they're a great compliment to therapy. I think they're really useful as a tool. I am not a huge fan of the app-based therapies you know. Like text therapy and stuff like that, I think it's not a good fit for everybody of course. But I think the apps themselves as far as tools and teaching and reinforcing skills I think are really helpful. And I would imagine for teenagers, I am assuming most of them are on their phones quite a bit. So that might be kind of a really easy transition or way to incorporate into something they're already doing. I think it would be cool especially for teens if there was a way to have some sort of feedback or back and forth with their therapist so they could actually share what they're doing, that would be neat. I think there's a lot of potential there for younger people.* [P14]

Although providers generally expressed support for mental health apps, most also raised several concerns. They described issues from both the provider and patient perspectives.

Theme 2: Cost/Access to Tool

The cost of purchasing an app was noted as a potential barrier. Providers suggested that they would only consider recommending apps with relatively low fees or a 1-time payment rather than ongoing costs.

> *Obviously free is best. Especially for teens, even young adults that are trying to make ends meet. If not then maybe something that's a one-time fee. If it was a subscription that would be harder for that population. I think that's something that older adults could possibly swing, but unless the parents are paying for it, it doesn't seem like that would be a good match.* [P13]

> *On the patient side yeah I think it would depend on how expensive the app would be. If they've got to pay $50/month that certainly could be prohibitive for folks.* [P5]

In addition, providers expressed concern that some youth, particularly those from low-income backgrounds, might lose phone service or have limited minutes, which could hinder access to the app and reduce engagement with its tools. They also noted that youth might have only intermittent access to mental health apps for various reasons, such as an inability to consistently pay for phone service or depleting their data or minutes. This loss of access could be frustrating for youth trying to use mental health apps and could interfere with their engagement with the app's content.

> *Most of our patients do have smart phones so that's usually not a barrier. But sometimes getting their cell phones shut off has happened before. So whether or not they'll have continuous access I think has been a barrier in the past.* [P8]

Theme 3: Engagement

Many providers expressed concern that a common barrier to using mental health apps is a lack of patient engagement. Several participants shared experiences of patients downloading apps but never using them beyond the initial session. Others noted that apps often lose their appeal quickly, making sustained engagement a challenge.

> *Yeah, and I think the literature is pretty clear that if you send patients off to use some sort of behavioral health app, they'll use it once or twice. They'll check it out and then use drops of precipitously.* [P1]

> *I think patients often download and forget about them which is probably the biggest one is the engagement with some of these apps is low. After a short amount of time they might forget about it and then they don't really come back to it. So in the context of therapy that was less of an issue because I would be there and be encouraging that use. So that's one of the barriers if people were just using it out in the real world.* [P9]

Providers described challenges in getting youth to engage with an app at all. Several noted concerns that apps must be visually appealing and up-to-date to attract teens, who are often highly tech-savvy. Based on their experience, if an app was not engaging enough, usage would decline, similar to what they had observed with other patient-focused apps.

> *I think we all know that teens are very sophisticated with apps and tech so I think having it be interesting and up to date is I think if they're going to use it, it's got to be flashy in some way.* [P3]

Theme 4: Patient Privacy

Most providers expressed concern about patient privacy. Some noted that their patients had raised concerns about the privacy of data collected by apps, which could make them hesitant to use mental health apps. Providers also emphasized that health systems would need to ensure patient privacy before integrating an app into treatment.

> *I know so many patients just in general are very concerned about their privacy. So I think that would be at the top of my list is just knowing that I could have confidence in saying to this person this app has*

> *been vetted in these different ways. This is the security that it offers. Maybe it's even password protected when you sign in so that way if your mom picks up the phone and she can't get into that particular app. I feel like that's a deeply personal thing. And I would want to make sure that I am protecting my patient's confidentiality and I can speak to how this app does that.* [P2]

> *I think always a concern is privacy and making sure the data is secure and HIPAA compliant. Especially working in health systems we're always concerned about technology and the privacy of patient information.* [P9]

Some providers also discussed concerns about patient privacy at home. Before recommending an app, they wanted to be confident in their understanding of how patient data were being protected.

> *On the other hand, if their parents are checking their phone and they don't feel like they have privacy they might not be able to use it in the same way that they might have otherwise had they had more privacy.* [P8]

Providers' Opinions About the Use of the BritePath App

Overview

All of the themes providers discussed regarding mental health apps in general were also reported in relation to the BritePath app. For instance, providers expressed privacy concerns both broadly and specifically in the context of BritePath. After hearing a description of BritePath, all but one provider expressed interest in trying this type of app for patients with depression, suicidality, or both. In addition to the general themes identified for mental health apps, several additional themes emerged specific to BritePath.

Theme 1: May Make Visits More Efficient

Providers described several ways in which an app such as BritePath could make therapy visits more time-efficient. Some noted that BritePath could help patients gain insight into the relationship between their mood and the activities they engage in between visits.

> *I think such a common thing with younger teens is being able to reflect back on how their week went or how their time went. And often times having something that is kind of tracking in the moment gives so much more actual information of when there were any spikes in anxiety or in distress in some way to help be like oh I saw this happened on this day, what happened then. To help anchor them into being able to talk about some things. So I can see that being a real benefit.* [P15]

> *I think it's so important for people to be able to see visually what the distribution is of the moods throughout the week. And also to remember that they're not always in that low point. So having some check ins throughout and maybe not only on their really bad days, right, but check ins every day would be helpful probably so they could be able to track down only the downs but maybe the ups or the little bit better.* [P14]

BritePath was also seen as a tool for providers to gather information on patient moods and their use of techniques between in-person sessions. Providers noted that access to this information via the app could make sessions more focused and productive by reducing the need to collect it during visits and allowing for better session preparation. Additionally, real-time patient-reported data on the app were perceived as more complete and valid than recalled information shared during a visit.

> *If you had the time to review what someone did or didn't do since you last saw them before they even walked in the door...I've saved at least 10 minutes of the precious 45 that we have together. I can launch straight into hey your mood meter was great over the last week. Or I see you check in on your safety plan on Tuesday, tell me a little bit about that? I don't have to say so how did the last week go? Oh Tuesday was bad. Tell me more about Tuesday and then wait for them to bring up the fact that they had to use their safety plan on Tuesday. We can just get straight to it. And it helps...not only the time saver, but hopefully not having to deal with retrospective recall.* [P5]

> *I think too if we have a session with a patient and they're just in that moment trying to recall their week is not as impactful as if you can actually go back and see what they're feeling in any particular time and probably would really help with recall bias.* [P8]

Providers also discussed how the BritePath app could facilitate communication with youth and enhance their engagement in the treatment process.

> *I think...leveraging digital health tool...I think another place this can be helpful is if kids engage in it, it gives you more to talk about in the session. You know doing sessions with teenagers a lot of times it's pulling teeth to get them to say anything. At least this would provide them with one, another way to keep them engaged and two, more fodder for talking to someone that may not be interested in filling 45 minutes' worth of talking.* [P5]

Theme 2: Provides Access to Safety Plans

Most providers noted that a key benefit of BritePath was giving patients convenient and consistent access to their safety plans through the app. They believed that because most youth carry their phones, they would be more likely to have their safety plan readily available in a crisis—unlike a paper version, which could be easily misplaced.

> *I think the nice thing about it, is it's all in one place. It's easily accessible on our phone. If they're in distress at any given moment they can pull it out and it's all right there. That makes it much more likely that they're going to use their skills in the moment, versus us giving them a piece of paper or emailing*

> *them or whatever else technology that would be used in the past. So I think that's a huge advantage.* [P13]

> *So patients can have a safety plan in the EHR but if they can't access it. We print it out and who knows where they left it. This gives them something, most people are pretty good about not leaving without their phone. So, it's a way that's there, it's accessible.* [P1]

However, providers also had questions and concerns about the functionality of storing the safety plan on a youth's phone. Some wondered whether the safety plan could be easily integrated into the patient's electronic health record. Another concern was whether providers would be notified when a patient accessed or used their safety plan.

> *Would the provider be alerted if there was a change? For example, if someone were to use the program or use part of their safety plan would that then ping me as a provider? Or would I just need to go and check on the patient status every so often to know those sorts of things? Another thing I might be curious about is how it might communicate with my EHR, if there was a way to import the information or download, even the safety plan for example so that way I could then upload it into the patient's chart. I think that would be important for me too.* [P2]

Theme 3: Offers an Additional Way to Communicate With Patients

Providers noted that BritePath could serve as an alternative communication tool for patients. Some recounted experiences with teens who struggled to express themselves and suggested that BritePath could offer these youth another way to connect with their providers. One provider also noted that this feature could enhance engagement in treatment for some patients.

> *Obviously, someone who is busy, if they're actually busy or just having issues with talking with a clinician face to face, this is a way to avoid that stress of direct interpersonal interaction. So that's a positive. And adolescents have all sorts of reasons why they sometimes don't want to engage with care. This is a way for them to get some care even if they're not feeling like interacting with their clinician that week. That I see as a benefit. Asynchronous so it can be on their time rather than someone else's time.* [P12]

> *If you think about teenagers, they are much more comfortable with technology and sometimes that's a great place to start for therapy or treatment with them because they might not feel comfortable talking to a professional.* [P9]

Several providers valued BritePath's 2-way communication feature, noting that it could facilitate shared decision-making and provide real-time updates to the treatment plan.

> *I think there's a real value in that shared decision-making component that can go into an app like the one you're describing. And it also becomes an important way of sort of communicating.* [P1]

> *I think that two-way communication is what I think could be the most useful if it's going between the patient and the therapist and you're able to interact back and forth. So in terms of modifying treatment plan or knowing what is resonating with the person.* [P3]

Theme 4: Flexibility and Personalization of the BritePath App

A number of providers expressed enthusiasm for BritePath's flexibility and personalization features. Several noted that allowing youth to customize the app with photos, videos, motivational reminders, or prompts for coping strategies could help sustain engagement, increase usage frequency, and accommodate diverse patient needs.

> *I think that's definitely a bonus. Sometimes the out of the box app works just fine. But the more you can tailor it or personalize it to someone's specific needs, the more likely they are to use it, so I think that that could definitely be useful!* [P9]

> *I love that there can be general tools but also some really individualized personal things in there in too. I just love the idea of having names and photos of people, loved ones and pets and things like that.* [P14]

One provider highlighted BritePath's potential to empower youth by allowing them to personalize their treatment and safety plan.

> *The place a teen usually goes to for soothing support regulation is the phone. But it often feels like the phone is just riddled with pitfalls of things that could make things worse. Maybe don't go on Instagram and compare your life to other people. So the fact that there's something right next to it potentially on their phone for them to dive into instead of something that could be potentially more disregulating or harmful I think is fabulous. The more opportunity I think there is to customize what's happening in that app the more I'd be inclined to really empower my client to tap into a wise place, tap into wiser parts of themselves, have those places and parts load what's in the app. So then they have access to it in those harder moments. I could see that being a particularly empowering thing.* [P11]

The collaborative functionality of the app was also seen as a benefit, as it enables providers to modify the app's content in collaboration with patients over time as their needs change.

> *I think that two-way communication is what I think could be the most useful if it's going between the patient and the therapist and you're able to interact back and forth. So in terms of modifying treatment plan or knowing what is resonating with the person.* [P3]

Theme 5: Concerns About Integration of BritePath Into Clinical Workflow

Providers expressed additional concerns specific to BritePath, beyond those related to mental health apps in general. One

concern was its impact on clinical workflow. For instance, providers questioned how well the app would integrate with existing electronic medical record systems. They also expressed concerns about its ease of use for busy clinicians.

> *Another thing I might be curious about is how it might communicate with my EHR, if there was a way to import the information or download, even the safety plan for example so that way I could then upload it into the patient's chart. I think that would be important for me too.* [P2]

> *Do I have to have a certain log in? Do I see a list of patients that I have that are using it? How cumbersome is it to get into that? And what's the patient's understanding and expectations for me, either monitoring that data or accessing that data or is there even some sort of portal to do that? Or is it more that the patient brings in their phone with their smart phone app and sort of shows me what they've been up to?* [P1]

In some cases, these concerns amplified existing worries about mental health apps in general, such as what might happen if a patient experiences a crisis. Additionally, several providers raised concerns about managing patient expectations regarding 24/7 access to their provider and ensuring that patients understood how to handle crises when providers were unavailable.

> *The only thing that has come up before... as far as cell phone and texting technology is somebody trying to text you in crisis while you're off work. You're not checking, so if there's something that presents in the app that is alarming or suggests that this client is in danger of harming themselves and you haven't checked the app in whatever amount of time, that would be I guess the concern for me too.* [P13]

> *I would want to make sure that whatever the app is that it stays within the boundaries of how I am available to the client or not. And I'd want it to be clear, it's tricky because if I can get a report about where they're at conceivably I could get a report when they're having a moment of increased distress and does that require me to then reach out to them too? So, I'd want it set up to be like hey either this app only releases this data to me X amount of times, so if you put in that you're having a hard time I am not going to see it.* [P11]

Discussion

Principal Findings

The focus of this qualitative study was to gain a deeper understanding of provider opinions on the use of mental health apps in general and the BritePath app specifically for treating teen and young adult patients with depression, suicidality, or both. Providers in this study had no prior exposure to the BritePath app. Most were aware of mental health apps in general and expressed interest in trying the BritePath app with patients with depression, suicidality, or both. Nearly all providers viewed mental health apps as a complement to other mental health treatments rather than a replacement for psychotherapy visits.

Mental Health Apps in General

Most providers were supportive of mental health apps, recognizing their potential to provide access to information and skills that could improve youth mental health. Key benefits included their availability at any time—particularly when providers were not accessible, such as late at night—and their ability to reinforce skills such as stress reduction between sessions. However, providers differed in their level of comfort with these tools. Some frequently recommended resources such as online cognitive behavioral therapy programs or anxiety management apps, while others were less enthusiastic. Most providers noted that mental health apps might be particularly useful for younger individuals, as they frequently use phones and other digital devices. These findings align with existing research on provider and patient perspectives, which indicate broad support for digital mental health tools [13,17,33-35]. However, most providers emphasized that apps alone were not sufficient for youth with significant mental health concerns. Many explicitly stated that these tools should complement, rather than replace, psychotherapy.

All providers expressed concerns about patient privacy. They discussed potential risks related to data collection by apps, including unauthorized access by individuals other than the youth (eg, parents) and uncertainties about how app providers might use collected data. While privacy concerns are common with app usage in general [13], providers emphasized that data collected by mental health apps are particularly sensitive and require high standards of privacy and protection. Several providers noted that app usage has become so commonplace that privacy policies are often overlooked, which may be especially problematic for youth. One provider specifically highlighted the difficulty in determining how commercial apps safeguard data privacy, which they found concerning.

Many providers also raised concerns about maintaining youth engagement with apps. Difficulty in sustaining engagement with digital tools has been noted in prior studies [36,37]. Several providers pointed out that they currently have little to no insight into how engaged their patients are with mental health apps. This uncertainty tempered some providers' optimism about the potential of these tools as a strong aid in treatment.

Providers also identified several logistical barriers to the effective use of mental health apps. The cost was a common concern, particularly the potential inaccessibility of apps that require ongoing payments (eg, monthly fees). Additionally, several providers noted that youth often experience inconsistent access to phones due to service suspensions, late payments, or lack of internet access. While these concerns have been raised in studies on app use in general [13], they may be especially critical for youth with mental health concerns, as intermittent access could disrupt their ability to use these tools as a coping mechanism.

BritePath App

Each of the potential benefits and concerns providers expressed about mental health apps in general were also raised in relation

to BritePath, including data privacy, patient engagement, and cost-related concerns. In some cases, providers expressed stronger opinions—both positive and negative—specifically regarding BritePath. For example, concerns about patient privacy were heightened due to the app's inclusion of sensitive information on symptoms and suicidality. By contrast, providers strongly supported BritePath's potential to enhance psychotherapy visits by facilitating more in-depth collaboration between providers and patients on skill development.

Providers raised several additional themes regarding BritePath. Many highlighted the value of the app's personalization features, noting that these could enhance patient engagement. Most providers expressed enthusiasm for BritePath's ability to facilitate communication between sessions and help providers prepare more effectively for upcoming appointments.

Providers also raised concerns about ensuring that safety issues were thoroughly addressed if BritePath were to be used in practice. While they appreciated the idea of having a safety plan integrated into the app, they emphasized the need for built-in safeguards, such as pop-up messages directing patients to emergency services when necessary. Additionally, providers wanted to ensure that patients understood that communication through the app would not guarantee a 24/7 response from providers.

Practical Implications for Research and Clinical Practice

Providers suggested several ideas for future research on apps addressing suicidality in youth. Integrating ambient phone data, such as activity tracking, could enable more timely and accurate recording of safety plan activities, reducing reliance on self-reporting. Additionally, visually linking mood and activity data, along with incorporating pop-up mood assessments, could enhance patient awareness of the relationship between mood and behavior. Increasing personalization in mood tracking and allowing patients to annotate their mood data may further improve engagement with the app.

Although providers were enthusiastic about the potential benefits of a customizable, readily accessible safety plan on a patient's phone, many also raised concerns about ensuring the security of this information. From a clinical perspective, safeguarding safety plans on the device would be essential for successful implementation within health systems. Additionally, embedding a clear and transparent patient consent process into the app—one that explains who will have access to their data and why—would be crucial in helping patients understand the risks associated with using a tool such as BritePath. Furthermore, ensuring that BritePath is compatible with different phone operating systems and can integrate with various electronic health record systems would be key to facilitating widespread adoption in clinical practice.

Strengths and Limitations

This study provides new insights into providers' perspectives on the use of mental health apps for youth at risk of suicidal behavior. We acknowledge the potential for sampling bias. To address this, we aimed to recruit a diverse range of providers from different health systems affiliated with the NIMH Mental Health Research Network [23]. While this sample may not be fully representative of providers across all health systems, participants were drawn from health systems in 7 US states. Although the sample is not fully representative of all providers in the United States, it includes a variety of geographic areas, organizational arrangements, and provider types, ensuring a diverse range of perspectives. In designing our qualitative interview guide, we focused on usability and concept testing rather than adhering to a specific technology adoption theory, such as the Technology Acceptance Model [26] or the Diffusion of Innovation Model [27]. However, our interview questions covered many of the key domains included in these models. We invited providers who delivered mental health services within these health systems to participate, without requiring a specific level of experience with apps or other digital tools in their practice, as we aimed to capture a broad range of experience levels. It is possible that providers with a greater interest in app use were more likely to participate.

One goal of this study was to explore providers' opinions about the BritePath app, which was developed for use with youth at risk for suicide. Specifically, we sought feedback from providers with no prior clinical experience or knowledge of BritePath to assess its potential for future implementation in health systems. Our focus was on understanding the broader applicability of mental health apps, with a particular emphasis on BritePath. While we collected limited demographic data on providers, it is possible that unmeasured factors influenced their perspectives on the use of mental health apps. For example, although respondents represented a range of age groups, most were younger providers, who may have been more comfortable with apps in general than older providers. Despite these limitations, this study offers new insights into providers' perspectives on mental health apps, including the BritePath app, and highlights key concerns regarding their use with young patients at risk for suicide.

Conclusions

The use of mental health apps is expanding, yet more research is needed to determine how they can be most effectively integrated into mental health treatment. Most providers expressed interest in using the BritePath app for patients with depression, suicidality, or both; however, concerns about privacy and safety remain.

Acknowledgments

This study was supported by the National Institute of Mental Health: The Center for Enhancing Triage and Utilization for Depression and Emergent Suicidality in Pediatric Primary Care (grant P50MH115838). The development and testing of BritePath have been funded by the National Institute of Mental Health (grants R34 MH100451 and P50MH115838), the American Foundation

for Suicide Prevention (AFSP), and a clinical use version institutional review board protocol (STUDY20050047) for collecting usability data. The authors thank Ms. Jill Pope for editorial assistance, Dr. Megan Hamm for feedback on the interview guide, Ms. Brandie George-Milford for editorial support, and Dr. David Brent for leading the P50 grant.

Conflicts of Interest

The authors have an SBIR proposal with Ksana Health to commercialize the BRITE app.

Multimedia Appendix 1
Qualitative interview guide.
[DOCX File , 3849 KB - humanfactors_v12i1e64867_app1.docx]

References

1. Van Meter AR, Knowles EA, Mintz EH. Systematic review and meta-analysis: international prevalence of suicidal ideation and attempt in youth. J Am Acad Child Adolesc Psychiatry 2023 Sep;62(9):973-986. [doi: 10.1016/j.jaac.2022.07.867] [Medline: 36563876]
2. Xiao Y, Cerel J, Mann JJ. Temporal trends in suicidal ideation and attempts among US adolescents by sex and race/ethnicity, 1991-2019. JAMA Netw Open 2021 Jun 01;4(6):e2113513 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.13513] [Medline: 34125218]
3. Madigan S, Korczak DJ, Vaillancourt T, Racine N, Hopkins WG, Pador P, et al. Comparison of paediatric emergency department visits for attempted suicide, self-harm, and suicidal ideation before and during the COVID-19 pandemic: a systematic review and meta-analysis. Lancet Psychiatry 2023 May;10(5):342-351 [FREE Full text] [doi: 10.1016/S2215-0366(23)00036-6] [Medline: 36907199]
4. Brewer AG, Doss W, Sheehan KM, Davis MM, Feinglass JM. Trends in suicidal ideation-related emergency department visits for youth in Illinois: 2016-2021. Pediatrics 2022 Dec 01;150(6):39-48. [doi: 10.1542/peds.2022-056793] [Medline: 36373281]
5. Bridge JA, Ruch DA, Sheftall AH, Hahm HC, O'Keefe VM, Fontanella CA, et al. Youth suicide during the first year of the COVID-19 pandemic. Pediatrics 2023 Mar 01;151(3):1-15 [FREE Full text] [doi: 10.1542/peds.2022-058375] [Medline: 36789551]
6. Lantos JD, Yeh H, Raza F, Connelly M, Goggin K, Sullivant SA. Suicide risk in adolescents during the COVID-19 pandemic. Pediatrics 2022 Feb 01;149(2):2-7. [doi: 10.1542/peds.2021-053486] [Medline: 34977942]
7. McCoy K, Kohlbeck S. Intersectionality in pandemic youth suicide attempt trends. Suicide Life Threat Behav 2022 Oct;52(5):983-993 [FREE Full text] [doi: 10.1111/sltb.12895] [Medline: 35735265]
8. Durante JC, Lau M. Adolescents, suicide, and the COVID-19 pandemic. Pediatr Ann 2022 Apr;51(4):e144-e149. [doi: 10.3928/19382359-20220317-02] [Medline: 35417315]
9. Brent DA, Goldstein TR, Benton TD. Bridging gaps in follow-up appointments after hospitalization and youth suicide: mental health care disparities matter. JAMA Netw Open 2020 Aug 03;3(8):1-3 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.13100] [Medline: 32780119]
10. Fontanella CA, Warner LA, Steelesmith DL, Brock G, Bridge JA, Campo JV. Association of timely outpatient mental health services for youths after psychiatric hospitalization with risk of death by suicide. JAMA Netw Open 2020 Aug 03;3(8):e2012887-e2012812 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.12887] [Medline: 32780122]
11. Doupnik SK, Rudd B, Schmutte T, Worsley D, Bowden CF, McCarthy E, et al. Association of suicide prevention interventions with subsequent suicide attempts, linkage to follow-up care, and depression symptoms for acute care settings. JAMA Psychiatry 2020 Oct 01;77(10):1021-1030. [doi: 10.1001/jamapsychiatry.2020.1586]
12. Blueprint for youth suicide prevention: strategies for clinical settings. American Academy of Pediatrics. 2021. URL: https://www.aap.org/suicideprevention [accessed 2023-05-31]
13. Van AM, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. Depress Anxiety 2017 Jun;34(6):526-539. [doi: 10.1002/da.22657] [Medline: 28569409]
14. Kao C, Liebovitz DM. Consumer mobile health apps: current state, barriers, and future directions. PM R 2017 May;9(5S):S106-S115. [doi: 10.1016/j.pmrj.2017.02.018] [Medline: 28527495]
15. Chan S, Li L, Torous J, Gratzer D, Yellowlees PM. Review of use of asynchronous technologies incorporated in mental health care. Curr Psychiatry Rep 2018 Aug 28;20(10):1-11. [doi: 10.1007/s11920-018-0954-3] [Medline: 30155593]
16. Martinengo L, Stona A, Griva K, Dazzan P, Pariante CM, von Wangenheim F, et al. Self-guided cognitive behavioral therapy apps for depression: systematic assessment of features, functionality, and congruence with evidence. J Med Internet Res 2021 Jul 30;23(7):1-17 [FREE Full text] [doi: 10.2196/27619] [Medline: 34328431]
17. Radovic A, Anderson N, Hamm M, George-Milford B, Fascetti C, Engster S, et al. Perceived facilitators of and barriers to implementation of a decision support tool for adolescent depression and suicidality screening: focus group and interview study. JMIR Ment Health 2021 Sep 15;8(9):1-16 [FREE Full text] [doi: 10.2196/26035] [Medline: 34524090]

18. Caldeira C, Chen Y, Chan L, Pham V, Chen Y, Zheng K. Mobile apps for mood tracking: an analysis of features and user reviews. AMIA Annu Symp Proc 2017;2017:495-504 [FREE Full text] [Medline: 29854114]
19. Marshall JM, Dunstan DA, Bartik W. Clinical or gimmickal: the use and effectiveness of mobile mental health apps for treating anxiety and depression. Aust N Z J Psychiatry 2020 Jan;54(1):20-28. [doi: 10.1177/0004867419876700] [Medline: 31552747]
20. Larsen M, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016 Apr 13;11(4):e0152285-e0152214 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
21. Kennard BD, Goldstein T, Foxwell AA, McMakin DL, Wolfe K, Biernesser C, et al. As safe as possible (ASAP): a brief app-supported inpatient intervention to prevent postdischarge suicidal behavior in hospitalized, suicidal adolescents. Am J Psychiatry 2018 Sep 01;175(9):864-872 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17101151] [Medline: 30021457]
22. Goldstein TR, Kennard BD, Porta G, Miller AO, Aguilar K, Bigley K, et al. Bridging gaps in care following hospitalization for suicidal adolescents: as safe as possible (ASAP) and BRITE app. J Am Acad Child Adolesc Psychiatry 2024 Jul 18:1-13 [FREE Full text] [doi: 10.1016/j.jaac.2024.06.008] [Medline: 39032815]
23. Mental Health Research Network. URL: https://mhresearchnetwork.org/accessed [accessed 2023-06-01]
24. Goodman LA. Snowball sampling. Ann Math Statist 1961 Mar;32(1):148-170. [doi: 10.1214/aoms/1177705148]
25. Palinkas LA, Horwitz SM, Green CA, Wisdom JP, Duan N, Hoagwood K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Adm Policy Ment Health 2015 Sep;42(5):533-544 [FREE Full text] [doi: 10.1007/s10488-013-0528-y] [Medline: 24193818]
26. 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):1-20 [FREE Full text] [doi: 10.2196/18072] [Medline: 32442132]
27. Warty RR, Smith V, Salih M, Fox D, Mcarthur SL, Mol BW, et al. Barriers to the diffusion of medical technologies within healthcare: a systematic review. IEEE Access 2021;9:139043-139058. [doi: 10.1109/access.2021.3118554]
28. Atlas.ti. 2020 Jan 31. URL: https://atlasti.com/ [accessed 2020-01-31]
29. Davis MM, Currey JM, Howk S, DeSordi MR, Boise L, Fagnan LJ, et al. A qualitative study of rural primary care clinician views on remote monitoring technologies. J Rural Health 2014;30(1):69-78 [FREE Full text] [doi: 10.1111/jrh.12027] [Medline: 24383486]
30. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
31. Joffe H. Thematic Analysis. In: Harper D, Thompson AR, editors. In Qualitative Research Methods in Mental HealthPsychotherapy: A Guide for StudentsPractitioners. Harper DThompson A. editors. Chichester: Wiley-Blackwell 2012, pp. Edinburgh, Scotland: Wiley-Blackwell; Jun 28, 2011:209-223.
32. Nicholas H, McDowall A. When work keeps us apart: a thematic analysis of the experience of business travellers. Community, Work & Family 2012 Aug;15(3):335-355. [doi: 10.1080/13668803.2012.668346]
33. Melia R, Monahan L, Duggan J, Bogue J, O'Sullivan M, Young K, et al. Exploring the experiences of mental health professionals engaged in the adoption of mobile health technology in Irish mental health services. BMC Psychiatry 2021 Aug 19;21(1):1-11 [FREE Full text] [doi: 10.1186/s12888-021-03426-5] [Medline: 34412601]
34. Silfee V, Williams K, Leber B, Kogan J, Nikolajski C, Szigethy E, et al. Health care provider perspectives on the use of a digital behavioral health app to support patients: qualitative study. JMIR Form Res 2021 Sep 28;5(9):1-12 [FREE Full text] [doi: 10.2196/28538] [Medline: 34529583]
35. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016 May 23;11(5):1-21 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
36. Amagai S, Pila S, Kaat AJ, Nowinski CJ, Gershon RC. Challenges in participant engagement and retention using mobile health apps: literature review. J Med Internet Res 2022 Apr 26;24(4):1-12 [FREE Full text] [doi: 10.2196/35120] [Medline: 35471414]
37. Williams JE, Pykett J. Mental health monitoring apps for depression and anxiety in children and young people: a scoping review and critical ecological analysis. Soc Sci Med 2022 Feb 12;297:1-15 [FREE Full text] [doi: 10.1016/j.socscimed.2022.114802] [Medline: 35192989]

Abbreviations

ETUDES Center: Center for Enhancing Treatment and Utilization for Depression and Emergent Suicidality
KPCHR: Kaiser Permanente Center for Health Research
NIMH: National Institute of Mental Health

Edited by H Witteman; submitted 12.08.24; peer-reviewed by R Melia, C Jacob; comments to author 04.11.24; revised version received 10.01.25; accepted 13.01.25; published 26.02.25.

Please cite as:
Lynch F, Cavese J, Fulton L, Vuckovic N, Brent D
Provider Perspectives on the Use of Mental Health Apps, and the BritePath App in Particular, With Adolescents at Risk for Suicidal Behavior: Qualitative Study
JMIR Hum Factors 2025;12:e64867
URL: https://humanfactors.jmir.org/2025/1/e64867
doi:10.2196/64867
PMID:

Review

Legitimacy as Social Infrastructure: A Critical Interpretive Synthesis of the Literature on Legitimacy in Health and Technology

Sydney Howe[1], MSc; Carin Uyl-de Groot[1], PhD; Rik Wehrens[1], PhD

Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

Corresponding Author:
Sydney Howe, MSc
Erasmus School of Health Policy and Management
Erasmus University Rotterdam
Burgemeester Oudlaan 50
Rotterdam, 3062 PA
The Netherlands
Phone: 31 107547119
Email: howe@eshpm.eur.nl

Abstract

Background: As technology is integrated into health care delivery, research on adoption and acceptance of health technologies leaves large gaps in practice and provides limited explanation of how and why certain technologies are adopted and others are not. In these discussions, the concept of legitimacy is omnipresent but often implicit and underdeveloped. There is no agreement about what legitimacy is or how it works across social science disciplines, despite a prolific volume of the literature centering legitimacy.

Objective: This study aims to explore the meaning of legitimacy in health and technology as conceptualized in the distinctive disciplines of organization and management studies, science and technology studies, and medical anthropology and sociology, including how legitimacy is produced and used. This allows us to critically combine insights across disciplines and generate new theory.

Methods: We conducted a critical interpretive synthesis literature review. Searches were conducted iteratively and were guided by preset eligibility criteria determined through thematic analysis, beginning with the selection of disciplines, followed by journals, and finally articles. We selected disciplines and journals in organization and management studies, science and technology studies, and medical anthropology and sociology using results from the Scopus and Web of Science databases and disciplinary expert–curated journal lists, focusing on the depth of legitimacy conceptualization. We selected 30 journals, yielding 796 abstracts.

Results: A total of 97 articles were included. The synthesis of the literature allowed us to produce a novel conceptualization of legitimacy as a form of social infrastructure, approaching legitimacy as a binding fabric of relationships, narratives, and materialities. We argue that the notion of legitimacy as social infrastructure is a flexible and adaptable framework for working with legitimacy both theoretically and practically.

Conclusions: The legitimacy as social infrastructure framework can aid both academics and decision makers by providing more coherent and holistic explanations for how and why new technologies are adopted or not in health care practice. For academics, our framework makes legitimacy and technology adoption empirically approachable from an ethnographic perspective; for decision makers, legitimacy as social infrastructure allows for a practical, action-oriented focus that can be assessed iteratively at any stage of the technology development and implementation process.

(JMIR Hum Factors 2025;12:e48955) doi:10.2196/48955

KEYWORDS

legitimacy; health technology; infrastructure; literature review; technology adoption; health care governance; technology acceptance; health care delivery; social infrastructure; critical interpretive synthesis

Introduction

Background

Technological innovations and health care delivery have become increasingly intertwined in recent decades [1,2]. With each promising innovation, new questions are raised about how to integrate the technology into health domains and the social implications of doing so [2-8]. This problem has given rise to evaluation processes such as health technology assessment [9] and a range of government institutions, such as the National Institute for Health and Care Excellence in the United Kingdom, to help make decisions about the appropriate use of technologies in various health care contexts [10].

Questions of legitimacy are implicit in debates about embedding technologies in health care, and these implicit questions about legitimacy permeate the literature about health and technology. While there is limited conceptual work on legitimacy in health and technology specifically, within the small body of work that exists, legitimacy is explored through a wide range of theoretical lenses and conceptualizations. However, there is no consensus about what legitimacy is or how it works nor is there much conversation among different branches of the literature. Across social science disciplines and medicine, legitimacy is affiliated with many concepts, including "acceptability" [11], "reasonableness" [12], and "transparency" [3,4]; these terms are sometimes conflated with legitimacy itself. Definitions of legitimacy range just as widely, especially among different disciplines: in organization and management studies (OMS), legitimacy is often defined as "normative appropriateness" [13,14], although science and technology studies (STS) and medical anthropology and sociology (MAS) include definitions such as "the acceptability of claims to authority" [15,16] and "accountability for reasonableness" [10] when the term is defined. Some disciplines, such as OMS, follow a strong tradition of explicit theorization of legitimacy, while others, such as medicine, usually do not explicitly problematize or define the word "legitimacy," although legitimacy and similar concepts may play a substantial role in discussions in the field. Within the same discipline, some authors use legitimacy interchangeably with other concepts [6], while others focus on a single subcategory of legitimacy developed from a narrow line of theoretical thought [17]. Given the breadth and depth of the literature on legitimacy in health and technology, it is clearly an important topic. A cross-disciplinary conceptualization of legitimacy that can be applied to questions about embedding technologies in health care would allow for a deeper understanding of these processes.

We chose to focus on legitimacy for several reasons. Legitimacy specifically is widely used but poorly defined within multiple fields studying health and technology and points toward social and organizational processes. Even though legitimacy is clearly important in studies of health and technology across many disciplines, it is abstract and there is very little cohesion both within and across social science disciplines around how it is produced and how it is used.

To understand the role of legitimacy in relation to the embedding and governance of technology in health care, we fleshed out the different ideas and underlying assumptions authors bring to the OMS, STS, and MAS literature. This paper explores three research questions: (1) What does legitimacy mean in the context of health and technology? (2) How is legitimacy produced? (3) How is legitimacy used to explain the challenges of embedding technologies in health care? On the basis of a synthesis of the literature, we developed a novel conceptualization of legitimacy as social infrastructure. This allowed us to focus on the aspects of legitimacy related to norms, materialities, semiotics, and specific relationships among stakeholders and larger systemic contexts.

Existing frameworks guiding policy intentions around technology adoption in health care often do not reflect the realities of everyday practice nor do they focus on the social processes of embedding technologies in health care [18]. Current models and frameworks for embedding technologies (in health care), such as the Unified Theory of Acceptance and Use of Technology [19] and the Technology Acceptance Model (TAM) [20,21], provide useful information about the judgments and perceptions of various actors in relation to a specific technology. While these models provide useful insights into actors' *intention* to adopt, as well as the ethical considerations they may perceive and institutional norms they may face, they do not often explain how and why certain technologies are or are not adopted or accepted *in practice* among particular groups and thus can provide only limited governance support [22]. Most technology adoption and acceptance models deal primarily with behavioral intention [20,21]. These models attempt to explain individual behavior intentions (as a proxy for actual behavior) through an analysis of a variety of factors, which can include beliefs, attitudes, perceived control, and perceived norms. This leaves substantial gaps, including the lack of attention to external factors impacting the embedding process [20] and any difference between behavioral intention and practice. Focusing on the impact of legitimacy within technology embedding processes can help fill these gaps by allowing for more comprehensive and nuanced conversations about technology adoption and acceptance.

A Cross-Disciplinary Legitimacy Conceptualization

The goal of this review is to produce a theoretical synthesis of how the legitimacy of health and technology is conceptualized. We aim to generate theoretical insights that will allow for a cross-disciplinary approach to legitimacy. To do this, we used the critical interpretive synthesis (CIS) review method to produce a cross-disciplinary understanding [23] of how legitimacy is produced and maintained without privileging one discipline's theoretical traditions over another's [24].

We used OMS, STS, and MAS as a disciplinary framework to understand the different strands of reasoning within legitimacy studies. OMS includes the most clearly developed theorization of legitimacy as a concept and several of the best-known review articles [14,25]. Legitimacy studies is a well-defined subarea of OMS research that explores legitimacy in the context of management, entrepreneurship, and institutional endeavors. OMS primarily focuses on how legitimacy works at the level of organizations and institutions, rather than among individual people or through technologies.

STS explores the social and cultural implications of science, technology, and technology development [26,27]. STS tends to focus on controversies, the unforeseen implications of technology development and adoption, and "black box" technologies [28] that inhibit full understanding. This contributes to a strong disciplinary interest in legitimacy, particularly of technology and medicine. STS centers interactions with technologies and technical change in legitimacy discussions, rather than separating the motivations of human actors from technologies and practices.

Although MAS developed from different traditions, they have partially converged in the study of medical cultures and interactions from a qualitative perspective [29], which is also the focus of legitimacy studies in MAS. MAS generally takes a skeptical view of taken-for-granted institutions. Many studies explore the impact of this taken-for-grantedness on people who are harmed by or left out of these institutions and systems. This leads to a natural questioning of legitimacy and deeper exploration of baseline assumptions, particularly in a health care context. Together, these disciplines explore health and health technologies with less focus on institutions and more focus on humans and connections among them. We hope that greater synthesis of different conceptualizations of legitimacy could provide a more holistic perspective about how and why technologies are or are not embedded in health care that integrates ideas from outside current frameworks based on adoption, acceptance, and implementation.

This review does not focus on ethics, although some articles that deal with ethics are included in this review. Within ethics, there is general consensus that legitimacy in health and technology is produced through ethical judgments and values-based reasoning [30-33]. While there is more cohesion in ethics than in other disciplines that deal with legitimacy, the scope of the literature is narrower and generally limited to moral or ethical legitimacy [3,34]. Legitimacy is often defined in ethical terms across several disciplines, including management, STS, and health policy. Articles with an ethical focus on health care technology adoption usually prioritize normative and moral dimensions of legitimacy, which may or may not be related to other aspects that contribute to overall legitimacy, such as utility or social acceptability [35]. While moral legitimacy is an important dimension of the subject, it has also already been studied and reviewed extensively and is far more clearly defined within the literature than other dimensions of legitimacy. Therefore, we chose to focus on other aspects of legitimacy to fill gaps in the existing social science literature. In addition, our review does not center legal approaches. Legal literature defines legitimacy according to principles of a participatory democracy [36]. Because the focus of this review is on social processes rather than democratic principles, we have chosen to center other means of legitimacy production and thus other branches of the literature, while acknowledging that democratic principles can play an important role [15,37].

Methods

Overview

In a research landscape that increasingly encourages and necessitates interdisciplinary collaboration, a CIS [24] allows us to develop a richer conceptual frame that can bring these different disciplines into conversation with one another [23,38]. The CIS method specifically addresses the limitations of conventional systematic reviews in cases where the objective is to generate a critical analysis of a body of the literature [24]. It draws on a tradition of qualitative inquiry: rather than summarizing the literature using preidentified key concepts, CIS uses an inductive, iterative process to interpret data, develop key concepts, and synthesize concepts with data through the process of the review itself [39]. Therefore, CIS produces a critical interpretation of the data reported in literature, which allows questioning of taken-for-granted assumptions [24].

We found CIS to be a particularly appropriate review style for an exploration of legitimacy in health and technology because, unlike traditional review styles, CIS is adept at dealing with "a large, amorphous, and complex body of literature" [24]. CIS synthesizes findings in a way that is both conceptually valuable to researchers and useful in informing policy. Furthermore, the CIS review style can manage the abstract nature of legitimacy across 3 disciplines without reducing the complexity of existing literature.

CIS is methodological and rigorously structured. However, it is also labor intensive compared to other review styles and deprioritizes transparency and full replicability available through systematic reviews. As in other kinds of interpretive research, different researchers using the same materials may have come to different conclusions, and full transparency of analysis cannot be expected at the same level as quantitative studies. We chose the CIS method in service of greater interpretive flexibility and conceptual depth and because it allowed us to account for the diversity of qualitative research without privileging a small range of studies that met specific and, here, less-relevant methodological standards.

Despite many empirical papers that address some aspects of legitimacy in relation to health and technology, there is a much more limited number of articles dealing with legitimacy in the domains of health and technology at a conceptual level. Therefore, we adopted broad definitions of both technology and health, including as many articles with strong legitimacy theorization as possible. "Health" in this paper refers to "human health and well-being"; "technology" includes "any devices, machinery, or equipment developed for the application of scientific knowledge that may impact human health." We acknowledge that there are potentially many other ways to operationalize these concepts that could be relevant to this review and health care in general [40]. Given the broad, conceptual focus of the review, as well as the wide range of distinctive technologies considered, we introduced 2 measures that allowed us to provide a clear focus. First, we zoomed in on conceptualizations of legitimacy; it was outside the scope to examine the specific empirical contexts and details of these conceptualizations (ie, the use of specific technologies within

specific health practices). Second, looking at specific disciplines proved helpful to demarcate boundaries around our topic. We used academic disciplines as the basis of our theoretical categories to give structure to a broad and abstract review, following in a strong disciplinary synthesis tradition established by other CIS researchers [23,38].

Our review was conducted in 4 cycles [38]. In the first cycle, we conducted an exploratory search in relevant journals with many articles on legitimacy in the context of health and technology. Using this search and subsequent iterations, we refined inclusion criteria (for instance, a threshold for legitimacy theorization) to determine a list of disciplines, journals within each discipline, and articles within these journals that would be a part of our review (see Multimedia Appendix 1 for details). In the second cycle, we read and coded included articles directly from the text, then grouped codes into thematic areas. Third, we analyzed legitimacy in health and technology within each discipline through interpretive disciplinary summaries. Finally, we developed a synthesizing argument across all 3 disciplines, based on these interpretive summaries.

Search Strategy

Journals within each discipline were selected for the review manually. Given the interdisciplinary nature of this review, we sought to include relevant journals within each discipline from a wide range to ensure diversity of perspectives.

The first author ran an initial search in Google Scholar and the Web of Science (WoS) database using the terms “legitimacy,” “health,” and “technology.” However, it became clear that regardless of how specific the search criteria became, most records returned through a database search would not include deep conceptualizations of the term “legitimacy.” This is because this term is commonly used across many disciplines as an unproblematic descriptor of technologies, legal processes, narrative choices, medical decisions, policies, and institutions. Therefore, we switched to a more iterative, manual strategy for discipline and subsequently journal selection (Textbox 1). The first author searched through a few of the major journals (based on impact factor or prior knowledge of the review team) in depth to understand whether studies addressing questions of legitimacy in a specific discipline generally included deeper conceptualization. This method allowed us to narrow down the disciplines to be included (see Multimedia Appendix 1 for details).

Textbox 1. Selected journals.

Organization and management studies

- Journal of Business Venturing
- Journal of Management Studies
- Organization Studies
- Organization Science
- Administrative Science Quarterly
- Health Care Management Review
- Strategic Entrepreneurship Journal
- Health Care Analysis
- Strategic Organization
- European Journal of Public Health

Science and technology studies

- Science as Culture
- Science, Technology and Human Values
- Social Studies of Science
- Social Epistemology
- Journal of Responsible Innovation
- AI & Society
- Big Data and Society
- Risk Analysis
- Information Communication & Society
- Research Policy

Medical anthropology and sociology

- Medical Anthropology
- Medical Anthropology Quarterly
- Anthropology & Medicine
- Qualitative Health Research
- Critical Public Health

  (interdisciplinary journal: some articles included within science and technology studies)
- Sociology of Health & Illness
- Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine
- Social Science and Medicine

  (interdisciplinary journal: some articles included within science and technology studies)
- Sociology-The Journal of the British Sociological Association
- Sociological Theory

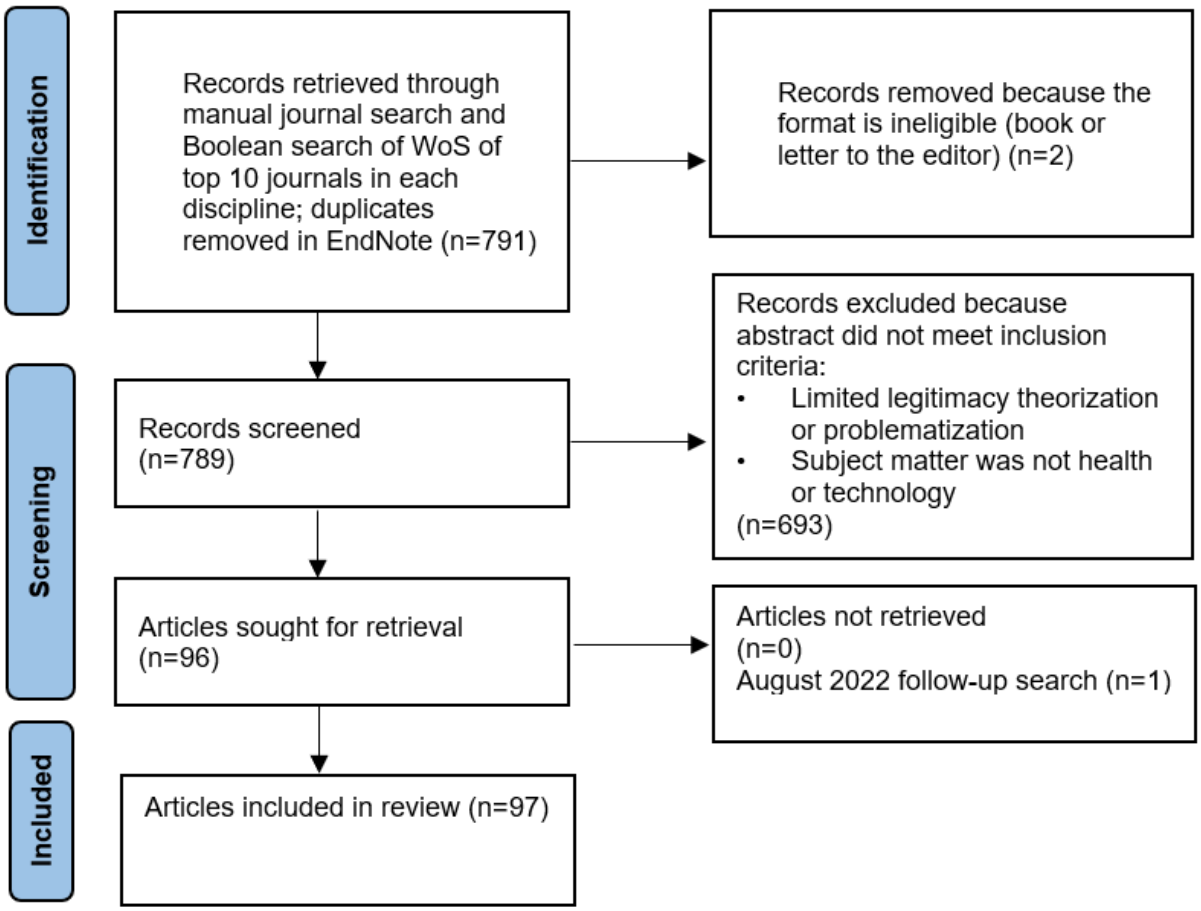
Upon the selection of disciplines, journals were first included based on impact factor: the top 10 journals in each discipline according to the WoS formed the base of the list. Then these journals were screened for topical relevance, and lower-ranked journals were excluded when the topic was not relevant. Finally, journals were added from library guides created by universities that are well-known in the field, especially if no definitive rankings were available from the WoS (STS) or there was too much disciplinary overlap in the rankings (MAS). This resulted in a list of approximately 20 journals per discipline. We then screened all journals for legitimacy, health, and technology results to produce the final list, prioritizing journals that contained significant theorization and conceptual work around legitimacy (see Multimedia Appendix 1 for details).

Selection Process

Overview

A total of 791 abstracts were selected for initial review in April 2021 (Figure 1). Articles were sorted based on inclusion criteria: an article was included if there was evidence of explicit theorization or problematization of legitimacy in the abstract and the subject of the article was health or technology or both. A final search for more recent articles was run in August 2022, resulting in 1 additional inclusion. Multimedia Appendix 1 gives more information about the inclusion criteria.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of inclusions. WoS: Web of Science.



Triangulation

The first author (SH) read all 791 abstracts and categorized them according to the inclusion criteria as "include," "exclude," or "review." The last author (RW reviewed 79/791, 10% of the total articles' abstracts at random (every 10th article when arranged in alphabetical order by author name). The resulting inclusion list closely matched SH's selection. In the 7 disagreements, the 2 reviewers came to a joint decision about each contested article and edited the inclusion criteria. SH then reviewed the abstracts again. With no disagreements following this protocol, the literature list of 96 articles was finalized in March 2021, with an addition of 1 article in August 2022, bringing the total to 97, including, by discipline, 48 for STS, 29 for OMS, and 20 for MAS [4-8,10,12,13,15,16,37,41-126].

Synthesis Methods

The goal of our initial analysis was to develop "synthetizing constructs" for each discipline, which "build on the explanations and interpretations of the constituent studies and are simultaneously consistent with the original results while extending beyond them" [24]. To accomplish this, SH first coded themes across a random selection of 30 articles spread equally across the 3 disciplines in Atlas.ti (ATLAS.ti Scientific Software Development GmbH). Inductive coding began with words taken from the article text. We made an exception for deductive codes relating to assumptions made as part of our "critical and reflexive approach" to this review [24]; for instance, coding whether articles discussed legitimacy as a process, perception, or asset based on our background understanding of existing legitimacy theorization across multiple disciplines. SH grouped codes by theme within each discipline. This thematic analysis was performed with the help of tables of code co-occurrences, generated in Atlas.ti, which showed overlap of unique codes within each document; this allowed us to understand when codes could be combined into a theme. We also used network visualizations, which helped us understand how different aspects of legitimacy may relate to one another on a broader scale (see Multimedia Appendix 2 for an example of one of these visualizations). This process allowed reviewer SH to maintain a differentiated perspective on each discipline and acted as a memory aid for the reviewer given the volume of the literature. SH then wrote summaries of the approach to

legitimacy within each discipline, which were discussed with RW. SH coded the remaining articles using Excel (Microsoft Corporation) while editing the disciplinary summaries as necessary. Multimedia Appendix 1 provides an example.

SH developed a synthesizing argument that could connect constructs of legitimacy across all 3 disciplines. SH tested the fit of different conceptualizations with the disciplinary summaries and then explored the development of these metaphors with RW. After SH had read 90% (87/97) of the articles, no further edits or additions to the disciplinary summaries or synthesizing argument were necessary, likely indicating data saturation.

Results

Organization and Management Studies

What Does Legitimacy Mean?

The basis of legitimacy epistemology in OMS can be traced to the highly influential paper by Suchman [14]. His definition of legitimacy still forms the basis of most subsequent theorizing in this field: "Legitimacy is a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" [14].

This definition builds from and expands on Weber's [127] understanding of legitimate authority, among other conceptualizations [41], focusing on the application of Weberian legitimacy concepts such as "tradition, affectual attitudes, value, and legality" [127] within a "socially constructed system of norms, values, beliefs, and definitions" [14]. Bitektine and Haack [128] further expanded on the Weberian connection to Suchman [14] with a multilevel legitimacy theory that differentiates among legitimacy judgments at the individual and collective levels [42]. Because of strong theoretical cohesion in this discipline, authors draw generously from these theoretical traditions, and most legitimacy subtypes are variations of normative legitimacy. Normative legitimacy deals with the norms, systems, and underlying assumptions that result in an institution, technology, or system being considered legitimate, and is generated through systems and social processes that may not be made explicit until there is a legitimacy crisis [41].

This definition is notable for its understanding of legitimacy as a perception of actions based on socially constructed norms and implicit conceptualization of legitimacy as a taken-for-granted state. However, perhaps more important is what is left out: because legitimacy is "generalized," it is not dependent on materialities or physical space, nor does the definition of legitimacy change specific to relationships among institutions or individuals. Our critical interpretive analysis shows that these kinds of assumptions seem to help determine the lines of reasoning for most OMS legitimacy scholars included in this review.

Our analysis shows that OMS legitimacy studies constitute a cohesive body of literature in which theoretical contributions are made incrementally. For instance, most authors outline exactly how they are adding to the existing literature (usually based on the definition by Suchman [14]) in a specific way. For instance, articles often divide legitimacy into ever more granular categories such as "network legitimacy" [43] and "actional legitimacy" [44] or focus on a particular aspect of the definition by Suchman [14] to explore in more depth, such as "normative appropriateness" [129]. Therefore it is easy to trace the lineage of theory. However, because every article adheres to the same framework, there is less diversity of thought. In light of this, a division could be made between technology literature and health care literature. Technology literature, usually with an entrepreneurship bent, tends to focus on audience perception and innovation framing through discourse, by teaching entrepreneurs how to appeal through discourse to different kinds of funding audiences [45]. For example, entrepreneurs seeking crowd-based funding are advised to make "contribution claims" emphasizing identity and added value to the funding community, rather than institutional ties or the significance of potential scientific or social advancements that may legitimate the entrepreneur to other audiences [45]. Health care literature focuses mainly on institutional legitimacy, for instance, by examining the negotiations involved in academic health center mergers [46] or the role of physician executives in institutional contexts [47]. For example, Kitchener [46] emphasizes that even as new sources of legitimacy for an academic health center became necessary in the wake of both a merger and a changing environment, the aim was to create an "aura" of legitimacy within a new institutional logic; in other words, even though the source of legitimacy changed, it was still defined through institutions.

How Is Legitimacy Produced?

On the basis of the literature reviewed for this study, conceptualizations of legitimacy production in OMS fall into 3 categories. Legitimacy-as-process, or legitimation, is "an interactive process of social construction" [25], often with clearly delineated steps toward generating legitimacy [48,130]. Articles using this conceptualization tend to track change over time [13,49,50], for instance, by examining the legitimation journey of a single innovation over decades [50]. Legitimacy as perception or "a social judgment, an evaluation, a socio-cognitive construction" [25] is used most frequently in relation to audiences or the question "legitimate to whom?" usually when examining innovations within larger organizational contexts [45,51,52]. Finally, a legitimacy-as-property conceptualization, in which legitimacy is viewed as an asset to be gained, increased, or lost [25], emphasizes how an entity gains legitimacy rather than differences between legitimate or illegitimate entities; this conceptualization appears to play a smaller role in more recent work. Legitimacy as a process and legitimacy as a perception are often used simultaneously within the same paper; there is an acknowledgment that both perspectives can offer valuable insights. For instance, McKnight and Zietsma [53] describe the process through which new ventures can achieve "optimal distinctiveness" (a perception) through framing and collaboration strategies (a process). Similarly, Garud et al [54] discuss entrepreneurial storytelling as a legitimation process that relies on audience perceptions, thereby promoting an understanding of audience perceptions

as part of analyses of different narrative techniques, which are framed as processes of legitimation [53].

How Is Legitimacy Used?

Within the literature reviewed, we noted that legitimacy is often assumed in OMS to be necessary for resource acquisition and is mostly relevant in an institutional setting. Because resource acquisition is the *raison d'être* of business in capitalism, legitimacy is considered essential. All literature included in our review aims to explain legitimation journeys or to advise institutions and entrepreneurs on how to move toward legitimacy. The stated goal is often to advance legitimacy theory within OMS, rather than explore empirical evidence beyond specific case studies.

Our analysis shows that nearly all legitimacy problems in OMS are framed as problems of categorization: either an institution or innovation has been miscategorized by its intended audience, or it has not yet been categorized [51]. Because of this, there are several articles that deal with the paradox of legitimacy for disruptive innovations: the innovation must conform to a category to be seen as legitimate, but it will not be seen as necessary unless it also distinguishes itself within the category [45,53,54]. Articles often describe how entrepreneurs can strike this balance or analyze how an older innovation navigated the paradox. For instance, Ruef and Scott [41] produced a multidimensional model of legitimacy designed to help hospitals use knowledge of legitimacy norms and categorization to "attract managerial legitimacy." Other authors provide articles that, while academic, can also be interpreted as how-to guides for entrepreneurs attempting to formulate narratives for different audiences. Articles in this category posit that entrepreneurs must fit the narrative to the correct category of both the legitimacy problem [54] and the audience's perspective [45].

Articles discussing specific innovations often explore problems of categorization through "institutional logics" [45,46,55] or the forms normative legitimacy may take within a particular institution. These articles acknowledge that different institutions have different norms, and therefore, entrepreneurs and innovators must adapt their storytelling to "fit" with the institution. Our analysis of the OMS literature showed that institutions that are taken for granted have no narrative flexibility. This means that any changes made not only require buy-in from professionals but also immediate practical implementation of changes (such as shifting patient education responsibilities from physicians to nurses) to avoid becoming mired in default institutional processes [56]. However, an entrepreneurial venture wrestling with liability of newness [53] has nearly endless opportunities for narrative reframing and limited risk of stagnation in nonexistent institutional processes [45].

Summary

In short, OMS seems to heavily theorize granular definitions of legitimacy, derived mostly from normative legitimacy. Most articles describe how organizations can move from illegitimate to legitimate status. Legitimacy must be negotiated at multiple levels and can be conceptualized as a process, perception, and asset. On the basis of our initial analysis, legitimacy in OMS rests on taken-for-grantedness and categorization according to institutional logics; it is hampered by liability of newness [51] and difficulties fitting into an institutional logic or other category.

We now move on from the critical interpretation phase of CIS methodology and toward synthesis. We develop synthetic constructs to metaphorically unite major themes within each discipline. While the OMS line of reasoning is somewhat homogeneous, our interpretive analysis shows that there is acknowledgment that different institutions use different systems of thought, or institutional logics, to judge and negotiate legitimacy. Stakeholders must navigate these different systems by fitting themselves into clear categories. However, the system by which legitimacy is judged and negotiated is rarely fully navigable to any one stakeholder.

In this way, navigating legitimacy in OMS is like using a highway system. The roads themselves (in this case, pathways to legitimacy) are clear and concrete but can only be used by particular categories of vehicles in particular locations, in the same way that, for instance, entrepreneurs must use particular discourses (ie, roads) to legitimate their invention to certain audiences (locations) [45,50,54]. The metaphor of the highway system allows us to capture OMS' conceptualization of legitimacy as a rule-bound system of connections, driven by institutions. The highway system metaphor shows us that the OMS perspective on legitimacy is unusual for how fully theorized and comprehensive it is. However, in the same way that mapping a highway system alone does not describe everything that happens on the road, this metaphor also helps us articulate gaps in OMS theorization. OMS focuses on pathways to legitimacy, institutional involvement, and categorization; however, it can be difficult to explore the moments of legitimacy breakdown, unexpected contradictions within legitimacy journeys, and the influence of materialities on legitimacy within the concrete and bounded pathways that make up legitimacy conceptualization in OMS.

Science and Technology Studies

What Does Legitimacy Mean?

Within the legitimacy literature reviewed, nearly all STS articles dealt with health care or science practices on the edge of scientific validity or public acceptance, usually by focusing on controversial technologies. Authors seemed to pay significant attention to the audience for these controversies. Innovations explored within STS also include changes in health care and scientific systems or discourses, for instance, the process of creating new institutions within an academic environment [57].

In this way, it seems that legitimacy in STS can only be discussed easily when it is in question. This may be linked to STS's longstanding focus on controversies as strategic locations of research and the STS tendency to "stay with the trouble" [131]. Taken-for-grantedness may be considered such an integral part of legitimacy that it raises red flags in a discipline that is highly sensitive to unintended consequences.

However, in the STS literature reviewed, legitimacy itself is rarely the subject of an article. Our interpretive analysis shows that legitimacy is still undertheorized in STS, in that many

articles do not reference other legitimacy literature or define the term. When legitimacy is explored in depth, authors tend to deal with it as a type of knowledge coproduced at the site of controversy between human and nonhuman actors. For instance, 2 articles examine the demarcation of boundaries in discourse about future-facing technologies; in these articles, narratives about nanotechnology and forensic science both impact and are impacted by developments in the technology [16,58]. For instance, Selin [16] demonstrates that even if a single vision of the future of nanotechnology is accepted by scientists over all others, competing visions may still have an impact on the direction of technological development via other avenues, such as governance and funding. We interpret this as evidence that legitimacy is coproduced between the narrative around new scientific disciplines and concrete innovations within those disciplines.

A consequence of the focus on disputed boundaries that we observed in the STS literature seems to be that STS scholars treat legitimacy as mutable. Boundary work [59] articles conceptualize legitimacy as a process that happens through negotiation within relationships among people and institutions. Boundary work deals with how people and institutions negotiate changing roles, limits, and responsibilities, often in relation to knowledge practices [59]. Within our review, most authors seem to use boundary work only as an explanatory framework for discursive strategies, but some articles include writing and practices as instances of boundary work [60]. For instance, Granja and Machado [58] describe how forensic scientists clearly demarcate their expertise and the limits of their role in the justice system through both discursive strategies and practical performances of accountability such as fastidious record keeping as a strategy to maintain their legitimacy. While perception can play a role in this process, we found no STS boundary work articles within the literature reviewed that consider legitimacy to be merely an asset. We interpreted this as possible evidence of the STS disinclination to consider any relationship or object to be fixed and a general disciplinary focus on processes of knowledge production.

How Is Legitimacy Produced?

We found that processes of legitimacy coproduction of new technologies in health care are described in relationships among a variety of human and nonhuman actors [28]. STS seems to resist explanations that pin legitimacy's presence or absence on a single explanation or strategy. Instead, by examining the influence of nonhuman actors, STS encourages readers to embrace complexity in legitimacy relationships. For example, Barker [61] discusses how a drug marketing campaign legitimated a contested illness; McLevey [62] details how legitimacy goals for policy recommendations change how think tanks make knowledge; and Greco [63] uses the concept of "scientific ideology" to explore how the placebo effect threatens the legitimacy of scientific rationality in biomedicine. In these examples, the influence of unexpected nonhuman actors shapes legitimacy in a porous way; this influence adds complexity to legitimacy relationships. These examples thus turn the "expected" legitimacy relationship upside down. Barker shows that a treatment can make an illness legitimate (rather than an illness becoming legitimate and then treatment being developed). McLevey [62] demonstrates that knowledge may be made to legitimize policy recommendations (rather than knowledge first being created and then informing policy decisions). Finally, Greco [63] posits that a biomedical effect can threaten the legitimacy of biomedicine (rather than biomedicine creating biomedical effects that reinforce its legitimacy).

How Is Legitimacy Used?

Through our critical interpretive analysis, legitimacy in STS appears to be yet another "two-faced Janus" [28], in that it is coproduced by both taken-for-granted processes and unexpected assemblages, which may appear contradictory. For instance, Brown and Michael [64] note that the institutional mechanisms for validating scientific discovery play a part in legitimizing medical use of pig organs. However, the authors also find that these taken-for-granted institutional processes are not enough to legitimate transspecies transplantation. Rather, porcine donor organs are legitimated in their interpretation through an unexpected assemblage of institutional processes and scientific and cultural judgments about pigs. While anatomical similarities between humans' and pigs' organs exist from a scientific perspective, human-pig differences legitimate porcine donor organs (over, for instance, monkey organs) from a cultural perspective [64]. These apparent contradictions do not threaten the legitimacy of porcine donor organs; instead, the authors interpret these contradictions as necessary to establish legitimacy for a controversial innovation in medicine.

While most articles in the reviewed STS literature use the concept of legitimacy to explore complexity in relationships among technologies, ideas, people, and institutions, a small subset of the literature consists of discourse analyses focusing on language as the only vehicle of legitimation [4,62]. In these articles, there is no room for other means of legitimation, such as nondiscursive practices, language-free materialities, or physical space. In addition, policy-oriented articles about discourse-based legitimation strategies tend to view these strategies as mutually exclusive: "legitimacy can only be achieved if complexity is reduced via language" [65]. This appears to be a radical departure from the assembled view of legitimation present in most STS articles, in that legitimacy is produced by specific arguments designed to convince one party that another is legitimate, rather than *through* the complexity of negotiation among people, technologies, ideas, and institutions. However, both STS subgroups use legitimacy as a relational concept, whether relationships are implied through arguments or complex negotiations.

Summary

Our analysis demonstrates that STS approaches legitimacy through controversy and contestation. Because of its empirical orientation and wide variety, the way legitimacy is approached in STS cannot be viewed as theoretically cohesive. We found that STS explores legitimacy as an artifact of knowledge coproduction, centering nonhuman actors and complex legitimacy processes through discourse, practices, or both. In one conceptualization, legitimacy is viewed as a constant and complex negotiation among actors and their contexts. However, discourse analyses often view the complexity of these relationships as a problem for legitimacy rather than an integral

aspect of the concept. Despite this divergence, nearly all STS literature reviewed seems to assume that legitimacy is relational and includes nonhuman actors, contexts, and practices of knowledge production.

Moving toward our synthetic constructs, we posit that in STS, the coproduction of legitimacy through both taken-for-granted processes and unexpected assemblages functions such as the mutable infrastructure of a computer operating system. The hardware of the computer does not change (much), even though the overall capacity of the computer relies on constant updates of existing software. New operating systems build on what came before while adding something new to stay relevant. Similarly, the legitimacy of technologies in health care is viewed predominantly as being built on underlying infrastructures, such as evidence-based medicine (EBM) and laboratory protocols. However, it is also considered to be shaped by unexpected influences, such as the coexistence of traditional medicines with EBM standards [66] and popular conceptualizations of new technologies that do not align with the use of those technologies by experts [58]. Furthermore, legitimization of a technology may have unexpected consequences, such as a medicine legitimizing an illness [61]. Legitimacy, according to STS scholars, must be constantly negotiated among nonhuman actors through an iterative process that is most visible at the boundaries. Through the metaphor of computer operating systems, STS emphasizes a dualistic perspective of legitimacy. Legitimacy in STS depends on the mutability and unexpected creativity of practices, actors, and context (ie, software); simultaneously, this "software" of legitimacy must be built on top of preexisting systems that are much more rarely changed in substantial ways (ie, hardware). Without these preexisting systems, the knowledge production practices delineated by STS scholars have nowhere to operate.

Medical Anthropology and Sociology

What Is the Meaning of Legitimacy?

On the basis of our critical interpretation of the MAS literature under review, we found that MAS tends to center relationships among people within larger systems and physical spaces, rather than the nature of the systems or institutions themselves. Similar to STS, the MAS literature reviewed usually discusses controversies rather than strategies of legitimation.

Because few articles use concrete definitions of legitimacy or conceptualizations that reference any other body of literature, even if they discuss legitimacy issues in depth using empirical evidence, we found legitimacy to be undertheorized within MAS as a discipline. Several articles use terms such as "medical legitimacy" [6] or "narratives of legitimation" without definitions [67]. Some define legitimacy only in relation to another concept [12,68,69]. More than any other discipline, MAS seems to veer toward legitimacy definitions that are limited by subject and narrow [59] in scope: for instance, "being legitimately on sick leave requires both a certified illness and the inability to perform one's work as a result of that illness" [70]. In our interpretation, this appears to allow for more empirical grounding but prevents conversation with other literature about legitimacy as a larger concept. Within the MAS literature reviewed, legitimacy is always relative, in that there is no universal legitimacy that works across all contexts.

However, there is still some cohesion visible among a subset of articles in the discipline. Several authors frame legitimacy as a form of "symbolic capital" in the Bourdieuvian sense [69,71,132] that can be used for resource acquisition. These articles and others also use boundary work [59] to describe the legitimacy of professional roles as a negotiation of boundaries mostly through discourse [72]. In combination with the disciplinary tendency to focus on the disenfranchised, these framings paint legitimacy as a power issue [37].

The literature reviewed seems to prioritize nuanced accounts of both social processes in contested situations and the spaces and materialities that shape the contestation of a technology. For instance, in an exploration of medical quackery in South Africa, Hornberger [68] shows that even when using medically useless technologies, providers of these technologies can construct legitimacy for their patients because of contextual elements, such as lower education levels, local beliefs, and problems accessing proven medical treatments. Spaces, materialities, and relativity appear to be especially important in articles about disconnects between policy and practice in professional settings [73]. For example, Perrotta and Geampana [7] explore how in vitro fertilization (IVF) professionals navigate the disconnect between the standards of EBM and the enthusiasm of patients for the latest technologies. In another account of IVF, Barnreuther [84] notes that because the research took place in a physician private home in India, as opposed to an institutional setting in a Western country, the resulting technical success was deemed illegitimate and went unrecognized in both medicine and research for decades. We find that these kinds of nuanced accounts present a deeper and more empirically grounded understanding of the meaning of legitimacy in context, despite a lack of theorization.

How Is Legitimacy Produced?

Our critical interpretation of the MAS literature emphasizes the intertwined nature of practices, discourses, social processes, and materialities in the production of legitimacy. Authors often seem to assume these are impossible to separate from each other without losing meaning. For instance, in a study of the use of restraint in a hospital psychiatry ward, McKeown et al [74] examine how boundaries of space (ie, office or patient area), identity (ie, health care professional or patient), and diagnosis work together to legitimize or delegitimize the use of force in context. In this way, legitimacy is defined, not merely influenced, by the environment in which it is generated [75].

When legitimacy is conceptualized as a power problem, categorization seems to solidify the standing of already-legitimate actors. Meanwhile, power appears to become less accessible for the illegitimate when professional boundaries are unclear [69,76-78] or impermeable to outsiders [8,79,80]. Common MAS topics include contested illnesses and attempts of patients to receive treatment [70,80,81] and role negotiations of nonmedical practitioners within the medical establishment [69,79,82-84]. In both cases, power imbalance is usually framed as the primary obstacle to legitimacy, which is necessary for a more egalitarian redistribution of power. Although some authors

argue it is possible to have some power without legitimacy [72,82,84], the physical location of institutions, people, and technologies still defines flows of resources, manifesting in, for instance, differential treatment of patients at a rural hospital that would be unlikely at an urban hospital due to differences in hospital conditions rather than patient deservingness [85]. On the basis of these examples, our interpretation is that MAS and OMS have some overlap in this area in that both disciplines seem to approach legitimacy as an issue of access to resources.

How Is Legitimacy Used?

MAS literature reviewed often examines how legitimacy can flow through the taken-for-granted artifacts of professionalism. These artifacts include prescriptions [78] and complicated-looking fake medical machines [68]. Because the object itself is imbued with legitimacy through specific qualities in the professional context (eg, an appearance of being "official" or "high tech"), artifacts in these articles seem to reinforce the legitimacy of the pharmacist writing the prescription or the practitioner using the machine. Material artifacts may also include larger systems such as old IT infrastructures or preexisting contracts that impede desired change within institutions [73], implicitly reinforcing the legitimacy of institutional systems. While these materialities impact potential changes in legitimacy, they are usually not the subject of debate in the MAS literature. Therefore, it seems that the legitimacy of the prescription-writing pharmacist may be contested, but the prescription itself usually is not [78].

Summary

Our critical interpretation of the MAS literature is that legitimacy in this discipline is a relative and context-dependent concept. Legitimacy is often understood to be intertwined with power relations, but authors disagree about the nature of this connection and its implications. Finally, MAS articles center sites of controversy in which the subject has already been deemed illegitimate by focusing on space, materialities, and relationships. MAS authors appear to use these lenses to question "default" narratives of legitimacy in favor of exploring how and why a technology, system, group, or experience came to be viewed as illegitimate.

Here we develop our final disciplinary synthetic construct. Our interpretive analysis found that in MAS, legitimacy is physical, material, and always relative. Therefore, we employ the metaphor of a building's plumbing: each part of the system, from the pipes to the faucets, performs a different function, which defines both their physical form and location. No two buildings can have identical plumbing systems, because no two buildings, even if constructed similarly, are situated in exactly the same location—thus, they must connect to the local environment in different ways. In much the same way, the materialities of health and technology, and how these materialities interact with spaces and actors, are considered to define the legitimacy of health and technology in MAS. A "quack" technology reads as legitimate instead of "fake" because of both its physical form and the spatial context it inhabits [68]; an innovation such as IVF in the "wrong" space may never be recognized as innovation at all [84]. In both cases, if we understand legitimacy to be an infrastructure that is dependent on both space and materialities, we can better connect the different underlying systems and elements that comprise legitimacy for health technologies. A sink connected to a sewer line requires multiple specific parts, arranged in a particular way, running through a particular location, to function. In the cases noted earlier, materials, the spaces in which they are developed, and the spaces in which they are used must be arranged in particular ways for technologies to be considered legitimate, whether the audience is the international scientific establishment [84] or residents of a rural South African community [68]. The metaphor of a plumbing system emphasizes the categorical importance of context and physical materials in the construction of legitimacy: in MAS, legitimacy cannot exist without these. Therefore, this metaphor also helps us understand why MAS generally avoids comprehensive definitions or theorization of legitimacy; other than a few basic building blocks, systems of legitimacy for MAS are essentially viewed as nontransferrable between contexts.

Synthesizing Argument: Legitimacy as Social Infrastructure

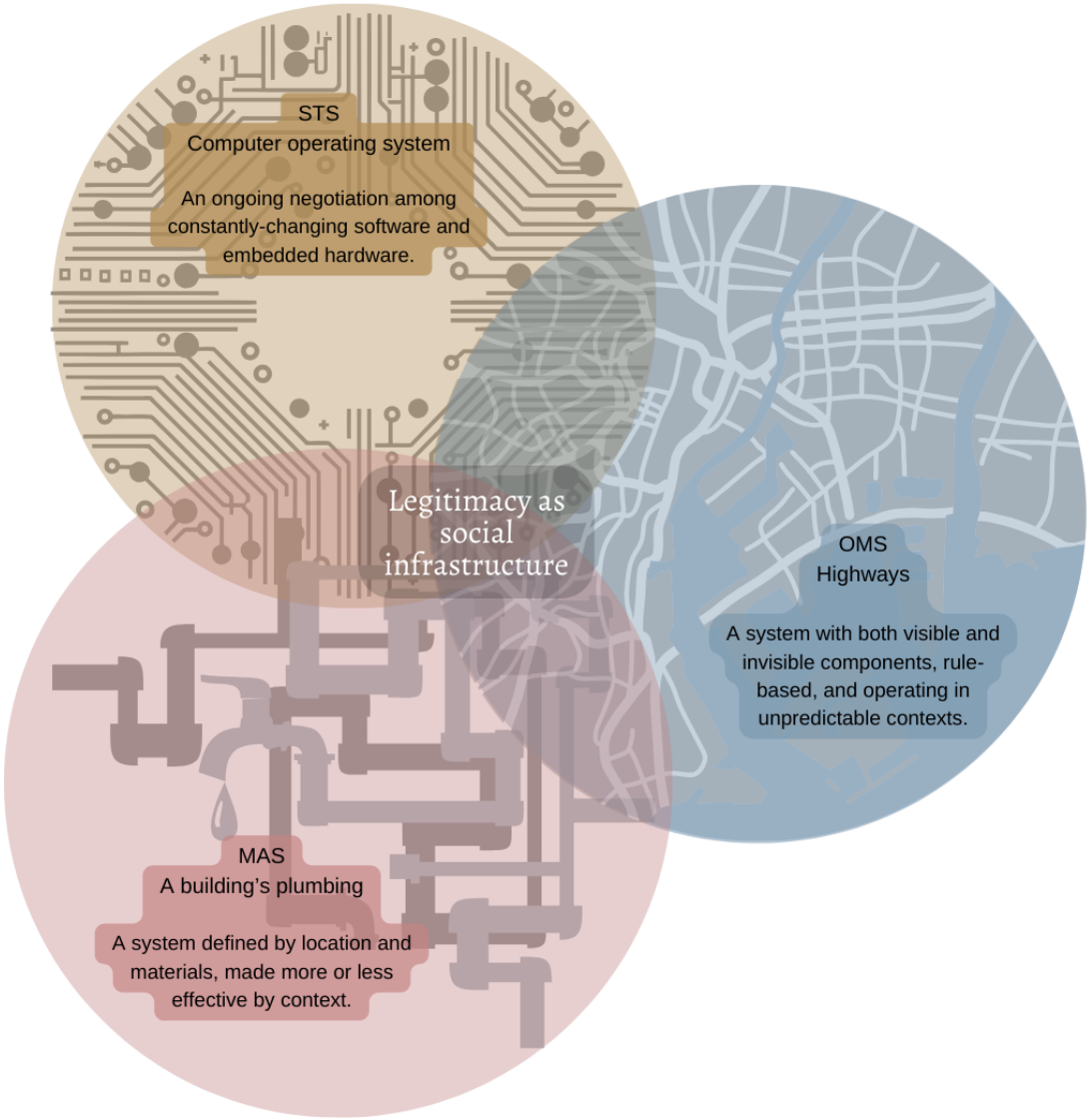
Legitimacy is omnipresent and important within all 3 disciplines. However, it is often undertheorized and the subject of many implicit assumptions across MAS and STS literature. In OMS, conversely, increasingly detailed theorization results in the proliferation of "subtypes" of legitimacy with limited empirical support (see Table 1 for more details). Our synthesizing argument allows us to bring these different disciplinary traditions into conversation with each other through a new conceptual framework for legitimacy.

Using the 3 disciplinary metaphors we developed, we propose that legitimacy can be conceptualized as a social infrastructure, based on the criteria for an infrastructure by Star and Ruhleder [133]. If we combine the metaphors of a highway system (OMS), computer operating system (STS), and plumbing system (MAS), we begin to see the building blocks of an infrastructure that can support an entire city (Figure 2). Each metaphor shows a particular element of city infrastructure (ie, understanding of legitimacy) clearly. However, similar to a layered map of a city's interconnecting infrastructures, we gain a fuller picture of how legitimacy works in health and technology when we layer the metaphors to view their intersections and divergences. Knowing how computer systems monitor traffic and manage stoplights helps us better understand both computer systems and highways; similarly, an STS understanding of how unexpected assemblages may work with taken-for-granted processes can help us better understand how these unexpected assemblages may impact theory about taken-for-granted processes in practice within OMS. Therefore, such a layered map allows for a richer conceptualization in which there is room for different approaches to legitimacy without denying the validity of any single conceptualization.

Table 1. Disciplinary characteristics: legitimacy in organization and management studies (OMS), science and technology studies (STS), and medical anthropology and sociology (MAS).

	OMS	STS	MAS
Theoretical foundations	Major focus on theory, granular definitions of legitimacy common; normative legitimacy most prevalent	Some theoretical exploration but limited theoretical cohesion	Limited theoretical exploration, reliance on granular definitions based on empirical evidence
Conceptualized as	A process (of negotiation); perception; or asset, and sometimes as several of these at once.	A product of knowledge coproduction; a negotiation among human and nonhuman actors, contexts, and epistemic cultures	A relative and context-dependent concept intertwined with power relationships
Purpose of literature	Describe how organizations can move from illegitimate to legitimate status	Investigate instances of controversy and contestation	Explore how and why a subject came to be viewed as illegitimate
Most important related concepts	Taken-for-grantedness; categorization and standardization	Relationships, taken-for-grantedness, and unexpected assemblages	Materialities, relationships, and space

Figure 2. Disciplinary metaphors. MAS: medical anthropology and sociology; OMS: organization and management studies; STS: science and technology studies.



Similar to a physical infrastructure, legitimacy can be understood as a binding fabric that makes possible certain types of relationships among people, institutions, and technologies. In our conceptualization, legitimacy is an inherently relational concept: it can only exist between two or more entities, rather than within the perceptions of a single person or institution. The notion of social infrastructure makes space for legitimacy to exist outside of human judgments, without devaluing the importance of perception in the production of legitimacy. Just as infrastructures consist of both materialities and abstract systems (eg, roads are made of concrete, but their use is facilitated by traffic laws), social infrastructure melds semiotic,

material, and normative aspects of legitimacy and explores how they might work together. Infrastructure thus allows for the integration of disparate and seemingly contradictory aspects of making legitimacy. While Star [134] has emphasized that infrastructure is inherently a "relational property," we use the word "social" to focus on how legitimacy supports specific types of relationships among people, institutions, and technologies; in particular, how these relationships are shaped by implicit and intangible norms and values [135]. For instance, understanding the nuances and norms of pharmacists' relationships with both physicians and health care institutions reveals why a simple policy change allowing pharmacists to prescribe medication legally is not enough to legitimize the practice [78]. This moves us away from conversations about legitimacy as a conduit for resources and toward more clarity about how legitimacy acts as a binding fabric within social relationships.

Descriptions of legitimacy across all 3 disciplines fit the criteria for an infrastructure by Star and Ruhleder [133]. For instance, legitimacy is usually only visible "upon breakdown" [133]. Contestation and controversy are the only sites of legitimacy exploration in MAS and STS; in OMS, legitimacy is discussed in terms of desired change in legitimacy status (ie, moving from illegitimacy to legitimacy), and contestation is captured explicitly through concepts such as "liability of newness" [53,54]. In addition, maintenance of this social infrastructure happens through continued adherence to legitimizing protocols and norms within relationships; deviations can result in a crisis of legitimacy. In this sense, legitimacy is characterized by a strong fit with predefined categories and "standards" [133] in many articles [45,53,65,85], even when legitimacy itself is not defined. In our conceptualization of legitimacy, norms and values make up the "installed base" into which legitimacy is embedded [133] and are responsible for legitimacy's taken-for-granted nature.

Legitimacy also acts similar to infrastructure, in that it can be used to invisibly support the work of organizations, technologies, or people [55,86,133], often by allowing for acquisition of resources or power [69,87]. In fact, values and norms are often so deeply entrenched and invisible that even the infrastructure supporting them (ie, legitimacy) becomes taken-for-granted and invisible itself. This combination can make it extremely difficult for any contradictory experiences or knowledge to find expression within a society's taken-for-granted, unquestioned truths [85,88,89,132]. Therefore, it is understandable why the moments of crisis when legitimacy is "visible upon breakdown" usually occur when norms and values clash with each other. For example, in the exploration of IVF by Perotta and Geampana [7], new reproductive technologies are a site of legitimacy crisis because they are heavily desired by patients but cannot adhere to standards of EBM. In this situation, physicians must navigate among the medical value of providing patients with the care they seek, the medical value of providing evidence-based care, and the medical norm of adhering to preset medical standards that may not have kept up with available care options.

Finally, the origins of legitimacy parallel those of any other infrastructure. Legitimacy is built on existing social systems: it "wrestles with the inertia of the installed base" [133] and borrows both strengths and weaknesses from that base, resulting in legitimation by once-functional mechanisms that are now relics. For instance, Sheard et al [73] show how desired changes in practices in a hospital setting are stalled by existing social, technical, and bureaucratic systems and entrenched social practices, despite a lack of effectiveness of these entrenched systems and a desire for change among the people acting within them. Overcoming this systemic inertia requires the integration of new practices and systems into the installed base, regardless of the current utility of that base; attempts to completely overhaul entrenched institutional practices or technologies without attention to the installed base rarely succeed.

In all 3 metaphors, infrastructure serves as a system through which different elements that contribute to legitimacy can negotiate with one another. A social infrastructure conceptualization reflects the nature and parameters of these negotiations. Thus this lens allows us to see where different perspectives on legitimacy could provide a more holistic picture of how legitimacy works in practice, encouraging a collaborative approach to conflicting views on legitimacy.

We hope this conceptualization can provide a useful lens to better understand how legitimacy is produced and how it works among institutions, technologies, and people. We aim to move the conversation toward a more holistic perspective on embedding technologies in health care that goes beyond models based primarily on judgments and perceptions.

Discussion

Principal Findings

Our conceptualization of legitimacy as social infrastructure allows for deeper analysis and understanding of the meaning, production, and uses of legitimacy when embedding technologies in health care. Our review contributes to the OMS literature by bringing in outside perspectives. Cross-fertilization among the 3 disciplines reigns in the proliferation of legitimacy theory in OMS while providing a better connection with empirical research. STS and MAS literature benefit from grounding empirical exploration within a more cohesive theoretical framework. Our review also helps contextualize studies of legitimacy that are more limited in scope; for instance, while discourse analyses of legitimacy are useful, this review demonstrates that discourse comprises just one aspect of legitimacy production [67,90,91]. By allowing for many definitions of legitimacy, the notion of legitimacy as social infrastructure is adaptable enough to produce meaning across disciplines while providing ways to explore relational, material, and semiotic aspects. By providing a framework through which to explore many conceptualizations and understandings of legitimacy, we hope LSI can facilitate fruitful cross-disciplinary collaborations of legitimacy researchers. Specifically, LSI allows different conceptualizations to coexist without competing with one another, providing a framework in which researchers can build on the work of other disciplines without having to disregard the traditions of their own discipline. In addition, this framework provides value to acceptance, adoption, and implementation researchers and other stakeholders interested

in better explaining how technologies are embedded in health care. Through an interdisciplinary understanding of legitimacy and a focus on relationships and moments of breakdown, LSI provides a mechanism to study technology embedding both contextually and over time. In this way, LSI produces a more comprehensive understanding of how the interconnected social, medical, environmental, and institutional elements surrounding a technology may interact in ways that impact legitimacy for a given technology, and thus, whether a technology will become embedded in health care.

Embedding Beyond Current Frameworks

Legitimacy as social infrastructure provides an alternative to current models describing how technologies are embedded in health care. We can use this conceptualization to account for gaps between policy intentions and practices when embedding technologies. Social infrastructure produces a more comprehensive understanding of the issues underpinning technology adoption than frameworks that hinge on trust [136,137], transparency [4,137], and acceptability [19], especially when those frameworks are based on behavioral intention [20,21,138]. Articles included in this review have shown that the complexities of embedding technologies in health care often clash with the best-intentioned policy choices because these frameworks generally do not account for the full scope of legitimizing processes within health care delivery systems. This is particularly evident in articles that focus on contested illnesses [80,89,92] and professional boundary work negotiations in light of new policies or collaborations in health care [47,56,77,78,93,94]. For example, Weiss and Sutton [78] demonstrate that even though policy changes explicitly allow pharmacists to prescribe medication, actually doing so in practice involves a more complex negotiation of roles and hierarchies to legitimize pharmacist prescribing activities. The new prescribing policy, based on notions of trust and acceptability, is not enough. For patients whose contested illnesses do not meet EBM criteria [80], negotiations and individual relationships are essential to access any kind of medical care [95]. Here, EBM criteria designed to increase transparency and trust in medicine prevent patients with real symptoms from accessing necessary care.

The many technology acceptance and adoption frameworks available currently have difficulty explaining adequately how and why some technologies are embedded in health care but others are not; in other words, “the literature does not specify the conditions for full use to be achieved” [138]. Similar to legitimacy, these terms are not clearly defined across the literature [138]: “acceptance” and “acceptability” generally are defined in terms of perceptions of potential users [138,139] and separate from both “adoption” and “appropriation,” which deal with the processes and practices of using the technology [138]. Other models aim to identify factors contributing to acceptance of new technologies. TAM is the most well-known example; it has been adapted and added to by numerous authors since its inception [20,21,138,140]. However, similar to most adoption and acceptance frameworks, TAM is based on behavioral intention, which has limited utility in predicting actual behavior [141,142] and does not generally attempt to integrate other systemic or relational aspects of technology acceptance. Within this review, Reay et al [56] demonstrated that gaining discursive buy-in, the focus of behavioral intention models of technology adoption, “is not enough” to legitimize new practices in a clinical setting without action and interpersonal, structural support.

Other frameworks have attempted to integrate sociocultural aspects of technology acceptance, specifically in low- and middle-income countries, although acceptance still generally focuses on individual users rather than systemic or relational factors [143]. More unusually, the nonadoption, abandonment, scale-up, spread, and sustainability framework by Greenhalgh et al [22], which is explicitly directed at health technologies, attempts to fill gaps left by TAM and other models by focusing on nonadoption and abandonment of technology instead of acceptance. This framework incorporates norms and values to some extent but does not focus specifically on legitimacy and the impact of norms and values on technology development.

The legitimacy as social infrastructure framework (LSI framework) is different from these models: rather than starting with adoption and acceptance of technologies, our conceptualization focuses on the relationships and negotiations that result in adoption and acceptance or not. If we take legitimacy as a necessary element in technology adoption and acceptance, conceptualizing legitimacy as social infrastructure allows us to understand how adoption, acceptance, and implementation frameworks may work together to explain the process of embedding technologies in health care. For instance, rather than conducting acceptance, adoption, and implementation studies separately in a health technology development project, researchers could use the conceptualization of legitimacy as social infrastructure as a framework to iteratively explore the binding fabric of relationships and legitimacy narratives that make up the landscape in which a new technology is being developed and provide advice to decision makers based on these findings. This focus would also allow project stakeholders to contextualize the results of a traditional acceptability or adoption study to better support the end goal of successfully embedding a technology in health care. Doing so would move conversation beyond simple adoption and acceptance and toward a more holistic perspective about this embedding process based instead on an analysis of legitimacy.

Focus on Relationships and Narratives

Momentum for embedding technologies in health care often fades due to unforeseen complexities [144] when moving technologies from laboratories to real-world environments. Particularly for mobile health technologies in which social and relational elements play a heightened role, the LSI framework could prove extremely useful in both anticipating and navigating complexity in the real world by describing changing relationships and narratives. Legitimacy as social infrastructure could provide a better blueprint that takes contextual factors into account earlier and avoids some of the acceptance and adoption bottlenecks that develop in linear models.

The struggles and unintended consequences of embedding technology in practice have been well-documented in STS and OMS [58,61,93,96]. The LSI framework conceptualizes how these documented struggles impact legitimacy and, thus,

technology adoption and acceptance. STS understands new technologies (including those designed for health care) to be both directional and open-ended [23]. Technologies are embodied with certain scripts and ideas, produced both through the development process and purposely in the minds of developers and other stakeholders, impacting how they can be used and when. However, the scripts managing this directionality never completely determine how a technology is used in practice [23,145]. Integrating technologies into real-world environments necessarily involves tinkering [22,23,144] by users, developers, and other stakeholders, often contradicting a technology's directionality entirely. For instance, as Perrotta and Geampana [7] demonstrate, new medical technology may be used in ways that contradict its original EBM-derived script, given a particular set of circumstances generated by patients, economics, and the pace of technology development. What happens next is a negotiation that determines how real-world context, practices, and technology scripts interact to generate a legitimacy relationship. Conceptualizing legitimacy as social infrastructure allows us to both understand the nature of these interactions and how they may or may not lead to technology adoption and acceptance in a health care setting. For instance, Sheard et al [73] discuss the complex systems and legitimacy relationships that impact the ability of care providers in the United Kingdom to act on patient feedback; this feedback is solicited in the first place to increase trust in and acceptability of the medical system. If we view patient feedback frameworks as a form of health care technology, legitimacy as social infrastructure helps us better understand and anticipate the relative importance of backgrounded relational and structural systems [73] that determined whether patient feedback was incorporated successfully in the health care institutional systems under study. An infrastructural understanding reveals how the 3 issues identified (ie, normative legitimacy, structural legitimacy, and organizational readiness [14,73]) work together to support or deter change in a complex organizational setting. A binding fabric of relationships ties all of these issues together and clarifies why the quality of certain relationships is more important than others. This focus also adds to understanding of the phenomenon of "institutional entrepreneurs." Relationships likely play a role in how institutional entrepreneurs avoid being "confined by the status quo," allowing them to change organizations from within despite legitimacy challenges for others [73].

Future Research and Applications

Given the utility of this conceptualization in bringing together 3 very different disciplinary traditions around legitimacy, future research could expand on this work to incorporate perspectives from disciplines, such as ethics and law, that we could not include here. In addition, it is possible that legitimacy is usually not explicitly conceptualized in medicine and psychology due to an implicit discipline-wide understanding of the values that underpin the concept. In other words, norms and legitimacy in medicine and psychology may be so closely intertwined with the goal of providing care that legitimacy may not be contested or questioned from these disciplinary perspectives. These kinds of taken-for-granted assumptions make it important to explore the origins, production, and meanings of legitimacy more thoroughly in the psychology and medical fields. Our conceptualization could allow for such investigation in fields in which the language describing legitimacy is either absent or perceived to be self-evident.

Potential applications of legitimacy as social infrastructure in practice have a wide range thanks to the flexibility and adaptability of the conceptualization. For example, to manage risks appropriately, regulators and others must have knowledge and understanding of the subject of regulation, in this case, health care technologies. This includes not only technical knowledge, but epistemological understanding of what is seen to constitute valid forms of knowledge, what is being left out, and "system knowledge of the context regulation is operating in" [146]. By approaching the embedding of technologies in health care as a result of knowledge coproduction, the notion of legitimacy as social infrastructure can provide deeper understandings of the systems and context of health technologies. It can then help regulators differentiate valid from invalid forms of knowledge in context. Regulators also experience legitimacy concerns themselves, in that regulation must be legitimate to function [9,147]. An understanding of legitimacy as social infrastructure could provide regulators with clarity about the practices, decisions, and structures that may impact their own legitimacy. For instance, knowing more about how their relationships with various stakeholders contribute to their legitimacy could help regulators take stakeholders' viewpoints into account at the most impactful moments and more appropriately narrate their decisions for different audience needs.

Limitations

Given the broad focus of our topic "legitimacy in health and technology," we included a heterogenous collection of empirical topics within each discipline. While this allowed for a broad discussion of legitimacy in health and technology, we did not produce one overarching definition of legitimacy; rather, we developed a conceptualization that can be used as a framework with the diversity of definitions already available. Similarly, the disciplinary boundaries are somewhat "fuzzy" [24]: OMS is a subdiscipline of sociology, STS overlaps with both MAS and OMS, and MAS is a combination of 2 disciplines with similar approaches. We found these disciplinary constructs useful because, despite their overlap, they highlight substantial differences among social science approaches to legitimacy.

In addition to the legal and ethical boundaries previously described, in this review, we did not engage with several other topics related to the legitimacy of technologies in health care. While the word "legitimacy" is used frequently in medicine and health economics, very few articles in these fields define the concept or explore it in any depth. Rather, the vast majority use the word uncritically within a surface explanation of the importance of a problem, either by claiming without evidence that legitimacy is an asset held by organizations, technologies, or people [148-150] or by mentioning that the legitimacy of a technology is under threat but without fully discussing why or how [151,152]. When legitimacy is defined, it is often reduced to a series of related concepts, such as "expertise, reciprocity, and trustworthiness," without discussing how these interact to

create legitimacy [153], although there are a few exceptions [9]. Due to the general lack of critical exploration, medicine and health economics were excluded as disciplines. This is an area that would benefit from further research using the LSI framework.

For similar reasons, we focused our search strategy within journals known to provide deep conceptualization of legitimacy (see Multimedia Appendix 1 for further details). Given the abstract and conceptual nature of our subject, a database search proved inadequate for finding all relevant articles; however, searching only within specific journals, despite rigorous selection, may have biased study findings. We feel the benefits of closely adhering to CIS protocol in this manner outweighed this issue due to our study's interpretive goals and the abstract subject matter.

Conclusions

The notion of legitimacy as social infrastructure provides a deeper understanding of how and why technologies are or are not embedded in health care. Social infrastructure focuses on the aspects of legitimacy related to norms, semiotics, materialities, and specific relationships among stakeholders and institutional contexts without excluding or privileging one form of knowledge production over the other. Through this conceptualization, the production and use of legitimacy can be studied ethnographically by following the social infrastructural networks and relationships that constitute and maintain legitimacy. By moving beyond judgment- and perception-based frameworks, legitimacy as social infrastructure approaches the problem of embedding technologies in health care through the lens of knowledge coproduction. The LSI framework enables empirical data collection and analysis that transcends disciplinary boundaries without negating the discrete lines of theoretical thought of any one discipline.

Given the interdisciplinary nature of health care policy and practice and the wide range of stakeholders involved, clearer and deeper communication around legitimacy could have far-reaching impacts. This could include new frameworks explaining technology adoption and acceptance, greater consensus about what makes a health care technology or policy decision legitimate, and strategic direction for decision makers working with disruptive health care technologies within entrenched systems.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Method details: selection of journals, disciplines, and articles.
[PDF File (Adobe PDF File), 234 KB - humanfactors_v12i1e48955_app1.pdf]

Multimedia Appendix 2
Methods: disciplinary network visualization.
[PDF File (Adobe PDF File), 529 KB - humanfactors_v12i1e48955_app2.pdf]

References

1. Ancker JS, Kern LM, Abramson E, Kaushal R. The Triangle Model for evaluating the effect of health information technology on healthcare quality and safety. J Am Med Inform Assoc 2012;19(1):61-65 [FREE Full text] [doi: 10.1136/amiajnl-2011-000385] [Medline: 21857023]
2. McDougall A, Goldszmidt M, Kinsella EA, Smith S, Lingard L. Collaboration and entanglement: an actor-network theory analysis of team-based intraprofessional care for patients with advanced heart failure. Soc Sci Med 2016 Sep;164:108-117 [FREE Full text] [doi: 10.1016/j.socscimed.2016.07.010] [Medline: 27490299]
3. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc 2016 Dec 01;3(2). [doi: 10.1177/2053951716679679]
4. de Fine Licht K, de Fine Licht J. Artificial intelligence, transparency, and public decision-making: why explanations are key when trying to produce perceived legitimacy. AI Soc 2020 Mar 19;35(4):917-926. [doi: 10.1007/s00146-020-00960-w]
5. Lehoux P, Williams-Jones B, Grimard D, Proulx S. Technologies of the self in public health: insights from public deliberations on cognitive and behavioural enhancement. Crit Public Health 2017 Mar 15;27(3):373-383. [doi: 10.1080/09581596.2017.1300637]
6. Fishman JR, Flatt MA, Settersten RA. Bioidentical hormones, menopausal women, and the lure of the "natural" in U.S. anti-aging medicine. Soc Sci Med 2015 May;132:79-87 [FREE Full text] [doi: 10.1016/j.socscimed.2015.02.027] [Medline: 25795991]
7. Perrotta M, Geampana A. The trouble with IVF and randomised control trials: professional legitimation narratives on time-lapse imaging and evidence-informed care. Soc Sci Med 2020 Aug;258:113115 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113115] [Medline: 32593117]

8. Petersen A, Tanner C, Munsie M. Between hope and evidence: how community advisors demarcate the boundary between legitimate and illegitimate stem cell treatments. Health (London) 2015 Mar 02;19(2):188-206. [doi: 10.1177/1363459314555240] [Medline: 25367895]
9. Cleemput I, Franken M, Koopmanschap M, le Polain M. European drug reimbursement systems' legitimacy: five-country comparison and policy tool. Int J Technol Assess Health Care 2012 Sep 17;28(4):358-366. [doi: 10.1017/s0266462312000529]
10. Russell J, Greenhalgh T. Affordability as a discursive accomplishment in a changing National Health Service. Soc Sci Med 2012 Dec;75(12):2463-2471. [doi: 10.1016/j.socscimed.2012.09.026] [Medline: 23103349]
11. Ward JK, Cafiero F, Fretigny R, Colgrove J, Seror V. France's citizen consultation on vaccination and the challenges of participatory democracy in health. Soc Sci Med 2019 Jan;220:73-80 [FREE Full text] [doi: 10.1016/j.socscimed.2018.10.032] [Medline: 30408684]
12. Wamsiedel M. Reasonableness: legitimate reasons for illegitimate presentations at the ED. Sociol Health Illn 2018 Nov 10;40(8):1347-1360. [doi: 10.1111/1467-9566.12776] [Medline: 29989187]
13. Haack P, Pfarrer MD, Scherer AG. Legitimacy - as - feeling: how affect leads to vertical legitimacy spillovers in transnational governance. J Manag Stud 2013 Dec 23;51(4):634-666. [doi: 10.1111/joms.12071]
14. Suchman MC. Managing legitimacy: strategic and institutional approaches. Acad Manag Rev 1995 Jun;20(3):571-610 [FREE Full text] [doi: 10.2307/258788]
15. Lövbrand E, Pielke RJ, Beck S. A democracy paradox in studies of science and technology. Sci Technol Hum Values 2010 Aug 26;36(4):474-496. [doi: 10.1177/0162243910366154]
16. Selin C. Expectations and the emergence of nanotechnology. Sci Technol Hum Values 2007 Mar 01;32(2):196-220. [doi: 10.1177/0162243906296918]
17. Fredriksson M, Tritter JQ. Disentangling patient and public involvement in healthcare decisions: why the difference matters. Sociol Health Illn 2017 Jan 11;39(1):95-111. [doi: 10.1111/1467-9566.12483] [Medline: 27862007]
18. Wehrens R, Sihag V, Sülz S, van Elten H, van Raaij E, de Bont A, et al. Understanding the uptake of big data in health care: protocol for a multinational mixed-methods study. JMIR Res Protoc 2020 Oct 22;9(10):e16779 [FREE Full text] [doi: 10.2196/16779] [Medline: 33090113]
19. Dwivedi YK, Shareef MA, Simintiras AC, Lal B, Weerakkody V. A generalised adoption model for services: a cross-country comparison of mobile health (m-health). Gov Inf Q 2016 Jan;33(1):174-187. [doi: 10.1016/j.giq.2015.06.003]
20. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007 Dec 05;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
21. Holden RJ, Karsh BT. The technology acceptance model: its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
22. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
23. Carboni C, Wehrens R, van der Veen R, de Bont A. Conceptualizing the digitalization of healthcare work: a metaphor-based Critical Interpretive Synthesis. Soc Sci Med 2022 Jan;292:114572 [FREE Full text] [doi: 10.1016/j.socscimed.2021.114572] [Medline: 34839086]
24. Dixon-Woods M, Cavers D, Agarwal S, Annandale E, Arthur A, Harvey J, et al. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. BMC Med Res Methodol 2006 Jul 26;6:35 [FREE Full text] [doi: 10.1186/1471-2288-6-35] [Medline: 16872487]
25. Suddaby R, Bitektine A, Haack P. Legitimacy. Acad Manag Ann 2017 Jan;11(1):451-478. [doi: 10.5465/annals.2015.0101]
26. Miller CA, Smith-Doerr L, Fouché R, Felt U. The Handbook of Science and Technology Studies. Cambridge, MA: MIT Press; 2017.
27. Woolgar S. The turn to technology in social studies of science. Sci Technol Hum Values 1991 Jan 01;16(1):20-50. [doi: 10.1177/016224399101600102]
28. Latour B. Science in Action: How to Follow Scientists and Engineers Through Society. Cambridge, MA: Harvard University Press; 1987.
29. Conrad P. Parallel play in medical anthropology and medical sociology. Am Soc 1997 Dec;28(4):90-100. [doi: 10.1007/s12108-997-1021-4]
30. Dawson A, Verweij M. Public health and legitimacy: or why there is still a place for substantive work in ethics. Public Health Ethics 2014 Jul 14;7(2):95-97. [doi: 10.1093/phe/phu018]
31. Tomlinson M, Chopra M, Hoosain N, Rudan I. A review of selected research priority setting processes at national level in low and middle income countries: towards fair and legitimate priority setting. Health Res Policy Syst 2011 May 15;9(1):19 [FREE Full text] [doi: 10.1186/1478-4505-9-19] [Medline: 21575144]
32. Boada JP, Maestre BR, Genís CT. The ethical issues of social assistive robotics: a critical literature review. Technol Soc 2021 Nov;67:101726. [doi: 10.1016/j.techsoc.2021.101726]
33. Crawford L. Moral legitimacy: the struggle of homeopathy in the NHS. Bioethics 2016 Feb 25;30(2):85-95 [FREE Full text] [doi: 10.1111/bioe.12227] [Medline: 26806451]

34. Morley J, Machado CC, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med 2020 Sep;260:113172. [doi: 10.1016/j.socscimed.2020.113172] [Medline: 32702587]
35. Baeroe K, Baltussen R. Legitimate healthcare limit setting in a real-world setting: integrating accountability for reasonableness and multi-criteria decision analysis. Public Health Ethics 2014 Mar 14;7(2):98-111. [doi: 10.1093/phe/phu006]
36. Schmidt VA. Democracy and legitimacy in the European Union revisited: input, output and 'throughput'. Political Stud 2012 Aug 06;61(1):2-22. [doi: 10.1111/j.1467-9248.2012.00962.x]
37. Martin GP. Representativeness, legitimacy and power in public involvement in health-service management. Soc Sci Med 2008 Dec;67(11):1757-1765. [doi: 10.1016/j.socscimed.2008.09.024] [Medline: 18922611]
38. Borst RA, Wehrens R, Bal R. Sustaining knowledge translation practices: a critical interpretive synthesis. Int J Health Policy Manag 2022 Dec 19;11(12):2793-2804 [FREE Full text] [doi: 10.34172/ijhpm.2022.6424] [Medline: 35279039]
39. Timmermans S, Tavory I. Theory construction in qualitative research: from grounded theory to abductive analysis. Soc Theory 2012 Sep 10;30(3):167-186. [doi: 10.1177/0735275112457914]
40. Berntsen GK, Gammon D, Steinsbekk A, Salamonsen A, Foss N, Ruland C, et al. How do we deal with multiple goals for care within an individual patient trajectory? A document content analysis of health service research papers on goals for care. BMJ Open 2015 Dec 10;5(12):e009403 [FREE Full text] [doi: 10.1136/bmjopen-2015-009403] [Medline: 26656243]
41. Ruef M, Scott WR. A multidimensional model of organizational legitimacy: hospital survival in changing institutional environments. Admin Sci Q 1998 Dec;43(4):877-904. [doi: 10.2307/2393619]
42. Haack P, Schilke O, Zucker L. Legitimacy revisited: disentangling propriety, validity, and consensus. J Manag Stud 2020 Aug 15;58(3):749-781. [doi: 10.1111/joms.12615]
43. Provan KG, Lamb G, Doyle M. Building legitimacy and the early growth of health networks for the uninsured. Health Care Manage Rev 2004;29(2):117-128. [doi: 10.1097/00004010-200404000-00005] [Medline: 15192984]
44. Wilmoth S. Foundation trusts and the problem of legitimacy. Health Care Anal 2004 Jun;12(2):157-169. [doi: 10.1023/b:hcan.0000041188.79809.05] [Medline: 15487816]
45. Fisher G, Kuratko DF, Bloodgood JM, Hornsby JS. Legitimate to whom? The challenge of audience diversity and new venture legitimacy. J Bus Ventur 2017 Jan;32(1):52-71. [doi: 10.1016/j.jbusvent.2016.10.005]
46. Kitchener M. Mobilizing the logic of managerialism in professional fields: the case of academic health centre mergers. Organ Stud 2002 May 01;23(3):391-420. [doi: 10.1177/0170840602233004]
47. Hoff TJ. The paradox of legitimacy: physician executives and the practice of medicine. Health Care Manage Rev 1999;24(4):54-64. [doi: 10.1097/00004010-199910000-00007] [Medline: 10572789]
48. Girschik V. Managing legitimacy in business‐driven social change: the role of relational work. J Manag Stud 2020 Jan 07;57(4):775-804. [doi: 10.1111/joms.12544]
49. Patala S, Korpivaara I, Jalkala A, Kuitunen A, Soppe B. Legitimacy under institutional change: how incumbents appropriate clean rhetoric for dirty technologies. Organ Stud 2017 Dec 17;40(3):395-419. [doi: 10.1177/0170840617736938]
50. Cattani G, Ferriani S, Lanza A. Deconstructing the outsider puzzle: the legitimation journey of novelty. Organ Stud 2017 Dec;28(6):965-992. [doi: 10.1287/orsc.2017.1161]
51. Alexy O, George G. Category divergence, straddling, and currency: open innovation and the legitimation of illegitimate categories. J Manag Stud 2012 Dec 18;50(2):173-203. [doi: 10.1111/joms.12000]
52. Golant BD, Sillince JA. The constitution of organizational legitimacy: a narrative perspective. Organ Stud 2007 Aug 01;28(8):1149-1167. [doi: 10.1177/0170840607075671]
53. McKnight B, Zietsma C. Finding the threshold: a configurational approach to optimal distinctiveness. J Bus Ventur 2018 Jul;33(4):493-512. [doi: 10.1016/j.jbusvent.2018.03.004]
54. Garud R, Schildt HA, Lant TK. Entrepreneurial storytelling, future expectations, and the paradox of legitimacy. Organ Sci 2014 Oct;25(5):1479-1492. [doi: 10.1287/orsc.2014.0915]
55. Patriotta G, Gond JP, Schultz F. Maintaining legitimacy: controversies, orders of worth, and public justifications. J Manag Stud 2011 Mar 17;48(8):1804-1836. [doi: 10.1111/j.1467-6486.2010.00990.x]
56. Reay T, Goodrick E, Casebeer A, Hinings CR. Legitimizing new practices in primary health care. Health Care Manage Rev 2013;38(1):9-19. [doi: 10.1097/HMR.0b013e31824501b6] [Medline: 22269782]
57. O'Kane C, Mangematin V, Geoghegan W, Fitzgerald C. University technology transfer offices: the search for identity to build legitimacy. Res Policy 2015 Mar;44(2):421-437. [doi: 10.1016/j.respol.2014.08.003]
58. Granja R, Machado H. Forensic DNA phenotyping and its politics of legitimation and contestation: views of forensic geneticists in Europe. Soc Stud Sci 2023 Dec 30;53(6):850-868 [FREE Full text] [doi: 10.1177/0306312720945033] [Medline: 32729409]
59. Gieryn TF. Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. Am Sociol Rev 1983 Dec;48(6):781-795 [FREE Full text]
60. Jacobson B. Constructing legitimacy in geoengineering discourse: the politics of representation in science policy literature. Sci Culture 2018 Apr 25;27(3):322-348. [doi: 10.1080/09505431.2018.1465910]
61. Barker KK. Listening to Lyrica: contested illnesses and pharmaceutical determinism. Soc Sci Med 2011 Sep;73(6):833-842. [doi: 10.1016/j.socscimed.2011.05.055] [Medline: 21840638]

62. McLevey J. Understanding policy research in liminal spaces: think tank responses to diverging principles of legitimacy. Soc Stud Sci 2015 Apr 30;45(2):270-293. [doi: 10.1177/0306312715575054] [Medline: 26477208]
63. Greco M. The ambivalence of error: "scientific ideology" in the history of the life sciences and psychosomatic medicine. Soc Sci Med 2004 Feb;58(4):687-696. [doi: 10.1016/s0277-9536(03)00220-x] [Medline: 14672585]
64. Brown N, Michael M. Switching between science and culture in transpecies transplantation. Sci Technol Hum Values 2001 Jan 01;26(1):3-22. [doi: 10.1177/016224390102600101] [Medline: 15362238]
65. Flink T, Kaldewey D. The new production of legitimacy: STI policy discourses beyond the contract metaphor. Res Policy 2018 Feb;47(1):14-22. [doi: 10.1016/j.respol.2017.09.008]
66. Cloatre E. Law and biomedicine and the making of 'genuine' traditional medicines in global health. Crit Public Health 2019 Jun 04;29(4):424-434 [FREE Full text] [doi: 10.1080/09581596.2019.1594696] [Medline: 31391708]
67. Foley L, Faircloth CA. Medicine as discursive resource: legitimation in the work narratives of midwives. Sociol Health Illn 2003 Mar;25(2):165-184 [FREE Full text] [doi: 10.1111/1467-9566.00330] [Medline: 14498938]
68. Hornberger J. Who is the fake one now? Questions of quackery, worldliness and legitimacy. Crit Public Health 2019 Jun 04;29(4):484-493. [doi: 10.1080/09581596.2019.1602719]
69. MacGregor C, Petersen A, Munsie M. From the margins to mainstream: how providers of autologous 'stem cell treatments' legitimise their practice in Australia. Health (London) 2021 Jan 13;25(1):51-68. [doi: 10.1177/1363459319846927] [Medline: 31081381]
70. Flinkfeldt M. 'Filling one's days': managing sick leave legitimacy in an online forum. Sociol Health Illn 2011 Jul 11;33(5):761-776 [FREE Full text] [doi: 10.1111/j.1467-9566.2011.01330.x] [Medline: 21561459]
71. Pushkar P, Tomkow L. Clinician-led evidence-based activism: a critical analysis. Crit Public Health 2020 Nov 27;31(2):235-244. [doi: 10.1080/09581596.2020.1841112]
72. Jones M, Pietilä I. Alignments and differentiations: people with illness experiences seeking legitimate positions as health service developers and producers. Health (London) 2020 May 16;24(3):223-240. [doi: 10.1177/1363459318800154] [Medline: 30222007]
73. Sheard L, Marsh C, O'Hara J, Armitage G, Wright J, Lawton R. The patient feedback response framework - understanding why UK hospital staff find it difficult to make improvements based on patient feedback: a qualitative study. Soc Sci Med 2017 Apr;178:19-27 [FREE Full text] [doi: 10.1016/j.socscimed.2017.02.005] [Medline: 28189820]
74. McKeown M, Thomson G, Scholes A, Jones F, Downe S, Price O, et al. Restraint minimisation in mental health care: legitimate or illegitimate force? An ethnographic study. Sociol Health Illn 2020 Mar 27;42(3):449-464 [FREE Full text] [doi: 10.1111/1467-9566.13015] [Medline: 31657030]
75. Busfield J. The moral economy of the pharmaceutical industry: legitimising prices. Health (London) 2021 May 17;25(3):271-287. [doi: 10.1177/1363459319879474] [Medline: 31623464]
76. Sanders T, Harrison S. Professional legitimacy claims in the multidisciplinary workplace: the case of heart failure care. Sociol Health Illn 2008 Mar 15;30(2):289-308 [FREE Full text] [doi: 10.1111/j.1467-9566.2007.01052.x] [Medline: 18290937]
77. Lees S, Palmer J, Procureur F, Blanchet K. Contested legitimacy for anthropologists involved in medical humanitarian action: experiences from the 2014-2016 West Africa Ebola epidemic. Anthropol Med 2020 Jun 04;27(2):125-143 [FREE Full text] [doi: 10.1080/13648470.2020.1742576] [Medline: 32363909]
78. Weiss MC, Sutton J. The changing nature of prescribing: pharmacists as prescribers and challenges to medical dominance. Sociol Health Illn 2009 Apr;31(3):406-421 [FREE Full text] [doi: 10.1111/j.1467-9566.2008.01142.x] [Medline: 19055585]
79. Mizrachi N. Epistemology and legitimacy in the production of anorexia nervosa in the journal Psychosomatic Medicine 1939–1979. Sociol Health Illn 2002 Jun 05;24(4):462-490. [doi: 10.1111/1467-9566.00304]
80. Zavestoski S, Brown P, McCormick S, Mayer B, D'Ottavi M, Lucove JC. Patient activism and the struggle for diagnosis: Gulf War illnesses and other medically unexplained physical symptoms in the US. Soc Sci Med 2004 Jan;58(1):161-175. [doi: 10.1016/s0277-9536(03)00157-6] [Medline: 14572929]
81. Mik-Meyer N, Obling AR. The negotiation of the sick role: general practitioners' classification of patients with medically unexplained symptoms. Sociol Health Illn 2012 Sep 05;34(7):1025-1038 [FREE Full text] [doi: 10.1111/j.1467-9566.2011.01448.x] [Medline: 22384857]
82. Nathan S, Stephenson N, Braithwaite J. Sidestepping questions of legitimacy: how community representatives manoeuvre to effect change in a health service. Health (London) 2014 Jan 30;18(1):23-40. [doi: 10.1177/1363459312473617] [Medline: 23364312]
83. Tovey P. Contingent legitimacy: U.K. alternative practitioners and inter-sectoral acceptance. Soc Sci Med 1997 Oct;45(7):1129-1133. [doi: 10.1016/s0277-9536(97)00010-5] [Medline: 9257404]
84. Bärnreuther S. Innovations 'out of place': controversies over IVF beginnings in India between 1978 and 2005. Med Anthropol 2016 Oct 05;35(1):73-89. [doi: 10.1080/01459740.2015.1094066] [Medline: 26436991]
85. Andersen HM. "Villagers": differential treatment in a Ghanaian hospital. Soc Sci Med 2004 Nov;59(10):2003-2012. [doi: 10.1016/j.socscimed.2004.03.005] [Medline: 15351468]
86. Latimer J. Commanding materials: (re)legitimating authority in the context of multi-disciplinary work. Sociology 2004 Oct 01;38(4):757-775. [doi: 10.1177/0038038504045863]

87. Kwak K, Yoon H. Unpacking transnational industry legitimacy dynamics, windows of opportunity, and latecomers' catch-up in complex product systems. Res Policy 2020 May;49(4):103954. [doi: 10.1016/j.respol.2020.103954]
88. Caiata-Zufferey M. Genetically at-risk status and individual agency. A qualitative study on asymptomatic women living with genetic risk of breast/ovarian cancer. Soc Sci Med 2015 May;132:141-148. [doi: 10.1016/j.socscimed.2015.03.037] [Medline: 25813728]
89. Paxman CG. "Everyone thinks I am just lazy": legitimacy narratives of Americans suffering from fibromyalgia. Health (London) 2021 Jan 20;25(1):121-137. [doi: 10.1177/1363459319857457] [Medline: 31216871]
90. Sapir A. Mythologizing the story of a scientific invention: constructing the legitimacy of research commercialization. Organ Stud 2019 Jan 28;41(6):799-820. [doi: 10.1177/0170840618814575]
91. Vaara E, Monin P. A recursive perspective on discursive legitimation and organizational action in mergers and acquisitions. Organ Sci 2010 Feb;21(1):3-22. [doi: 10.1287/orsc.1080.0394]
92. Clark S, Botterill LC. Contesting facts about wind farms in Australia and the legitimacy of adverse health effects. Health (London) 2018 Jul 01;22(4):337-355. [doi: 10.1177/1363459317693407] [Medline: 28401817]
93. Edwards KT. Methods of legitimation: how ethics committees decide which reasons count in public policy decision-making. Soc Sci Med 2014 Jul;113:34-41. [doi: 10.1016/j.socscimed.2014.04.043] [Medline: 24833251]
94. Vargas R. How health navigators legitimize the Affordable Care Act to the uninsured poor. Soc Sci Med 2016 Sep;165:263-270. [doi: 10.1016/j.socscimed.2016.01.012] [Medline: 26806315]
95. Haldar M, Engebretsen E, Album D. Legitimating the illegitimate: how doctors manage their knowledge of the prestige of diseases. Health (London) 2016 Nov;20(6):559-577. [doi: 10.1177/1363459315596798] [Medline: 26245482]
96. Kannan-Narasimhan R. Organizational ingenuity in nascent innovations: gaining resources and legitimacy through unconventional actions. Organ Stud 2014 Mar 05;35(4):483-509. [doi: 10.1177/0170840613517596]
97. Carr ES, Obertino-Norwood H. Legitimizing evidence: The trans-institutional life of evidence-based practice. Soc Sci Med 2022 Oct;310:115130. [doi: 10.1016/j.socscimed.2022.115130] [Medline: 36070634]
98. Albert M, Paradis E, Kuper A. Interdisciplinary promises versus practices in medicine: the decoupled experiences of social sciences and humanities scholars. Soc Sci Med 2015 Feb;126:17-25. [doi: 10.1016/j.socscimed.2014.12.004] [Medline: 25500163]
99. Alexiou K, Wiggins J. Measuring individual legitimacy perceptions: scale development and validation. Strateg Organ 2018 May 09;17(4):470-496. [doi: 10.1177/1476127018772862]
100. Andersen RS, Tørring ML, Vedsted P. Global health care-seeking discourses facing local clinical realities: exploring the case of cancer. Med Anthropol Q 2015 Jun 30;29(2):237-255. [doi: 10.1111/maq.12148] [Medline: 25355457]
101. Bartlett A, Lewis J, Reyes-Galindo L, Stephens N. The locus of legitimate interpretation in Big Data sciences: lessons for computational social science from -omic biology and high-energy physics. Big Data Soc 2018 May 08;5(1). [doi: 10.1177/2053951718768831]
102. Bensimon CM, Smith MJ, Pisartchik D, Sahni S, Upshur RE. The duty to care in an influenza pandemic: a qualitative study of Canadian public perspectives. Soc Sci Med 2012 Dec;75(12):2425-2430 [FREE Full text] [doi: 10.1016/j.socscimed.2012.09.021] [Medline: 23089615]
103. Carnabuci G, Operti E, Kovács B. The categorical imperative and structural reproduction: dynamics of technological entry in the semiconductor industry. Organ Sci 2015 Dec;26(6):1734-1751. [doi: 10.1287/orsc.2015.1011]
104. Dodworth K, Stewart E. Legitimating complementary therapies in the NHS: campaigning, care and epistemic labour. Health (London) 2022 Mar 07;26(2):244-262 [FREE Full text] [doi: 10.1177/1363459320931916] [Medline: 32508138]
105. Goodrick E, Reay T. Florence Nightingale endures: legitimizing a new professional role identity. J Manag Stud 2009 Nov 11;47(1):55-84. [doi: 10.1111/j.1467-6486.2009.00860.x]
106. Goodwin D, Pope C, Mort M, Smith A. Access, boundaries and their effects: legitimate participation in anaesthesia. Sociol Health Illn 2005 Sep 11;27(6):855-871 [FREE Full text] [doi: 10.1111/j.1467-9566.2005.00477.x] [Medline: 16283902]
107. Hillman A. 'Why must I wait?' The performance of legitimacy in a hospital emergency department. Sociol Health Illn 2014 May 20;36(4):485-499 [FREE Full text] [doi: 10.1111/1467-9566.12072] [Medline: 24053721]
108. Hirschkorn KA. Exclusive versus everyday forms of professional knowledge: legitimacy claims in conventional and alternative medicine. Sociol Health Illn 2006 Jul 03;28(5):533-557 [FREE Full text] [doi: 10.1111/j.1467-9566.2006.00506.x] [Medline: 16910946]
109. Kim PH, Croidieu G, Lippmann S. Responding from that vantage point: field position and discursive strategies of legitimation in the U.S. wireless telegraphy field. Organ Stud 2016 Jul 09;37(10):1417-1450. [doi: 10.1177/0170840616634132]
110. Kushida S, Kawashima M, Abe T. Why this clinic now? A context-sensitive aspect of accounting for visits. Soc Sci Med 2020 Nov;265:113278. [doi: 10.1016/j.socscimed.2020.113278] [Medline: 32947182]
111. Lê G. Trading legitimacy: everyday corruption and its consequences for medical regulation in southern Vietnam. Med Anthropol Q 2013 Sep 08;27(3):453-470. [doi: 10.1111/maq.12052] [Medline: 24248995]
112. Lehoux P, Daudelin G, Hivon M, Miller FA, Denis JL. How do values shape technology design? An exploration of what makes the pursuit of health and wealth legitimate in academic spin-offs. Sociol Health Illn 2014 Jun 21;36(5):738-755. [doi: 10.1111/1467-9566.12097] [Medline: 24444436]

113. Lewis M. De-legitimising complementary medicine: framings of the Friends of Science in Medicine-CAM debate in Australian media reports. Sociol Health Illn 2019 Jun 21;41(5):831-851. [doi: 10.1111/1467-9566.12865] [Medline: 30793332]
114. Lischka JA. Explicit terror prevention versus vague civil liberty: how the UK broadcasting news (de)legitimatise online mass surveillance since Edward Snowden's revelations. Inf Commun Soc 2016 Jul 18;20(5):665-682. [doi: 10.1080/1369118x.2016.1211721]
115. Lockett A, Currie G, Waring J, Finn R, Martin G. The role of institutional entrepreneurs in reforming healthcare. Soc Sci Med 2012 Feb;74(3):356-363. [doi: 10.1016/j.socscimed.2011.02.031] [Medline: 21440968]
116. Manigart S. The founding rate of venture capital firms in three European countries (1970–1990). J Bus Ventur 1994 Nov;9(6):525-541. [doi: 10.1016/0883-9026(94)90020-5]
117. Markard J, Wirth S, Truffer B. Institutional dynamics and technology legitimacy – a framework and a case study on biogas technology. Res Policy 2016 Feb;45(1):330-344. [doi: 10.1016/j.respol.2015.10.009]
118. McKendrick DG, Carroll GR. On the genesis of organizational forms: evidence from the market for disk arrays. Organ Sci 2001 Dec;12(6):661-682. [doi: 10.1287/orsc.12.6.661.10083]
119. Ozcan YA, Shukla RK, Tyler LH. Organizational performance in the community mental health care system: the need fulfillment perspective. Organ Sci 1997 Apr;8(2):176-191. [doi: 10.1287/orsc.8.2.176]
120. Petkova AP, Rindova VP, Gupta AK. No news is bad news: sensegiving activities, media attention, and venture capital funding of new technology organizations. Organ Sci 2013 Jun;24(3):865-888. [doi: 10.1287/orsc.1120.0759]
121. Ridge D, Kokanovic R, Broom A, Kirkpatrick S, Anderson C, Tanner C. "My dirty little habit": patient constructions of antidepressant use and the 'crisis' of legitimacy. Soc Sci Med 2015 Dec;146:53-61. [doi: 10.1016/j.socscimed.2015.10.012] [Medline: 26498732]
122. Stevenson FA, Britten N, Barry CA, Bardley CP, Barber N. Perceptions of legitimacy: the influence on medicine taking and prescribing. Health 2002;6(1):85-104. [doi: 10.1177/1363459302006001446]
123. Swartz A. Legacy, legitimacy, and possibility: an exploration of community health worker experience across the generations in Khayelitsha, South Africa. Med Anthropol Q 2013 Jun 26;27(2):139-154. [doi: 10.1111/maq.12020] [Medline: 23804283]
124. Uibu M. The emergence of new medical pluralism: the case study of Estonian medical doctor and spiritual teacher Luule Viilma. Anthropol Med 2021 Dec 28;28(4):445-460. [doi: 10.1080/13648470.2020.1785843] [Medline: 32854514]
125. van Dijk S, Berends H, Jelinek M, Romme AG, Weggeman M. Micro-institutional affordances and strategies of radical innovation. Organ Stud 2011 Nov 03;32(11):1485-1513. [doi: 10.1177/0170840611421253]
126. Wienroth M. Value beyond scientific validity: let's RULE (Reliability, Utility, LEgitimacy). J Responsible Innov 2020 Oct 30;7(sup1):92-103. [doi: 10.1080/23299460.2020.1835152]
127. Weber M. Economy and Society: A New Translation. Cambridge, MA: Harvard University Press; 2019.
128. Bitektine A, Haack P. The "macro" and the "micro" of legitimacy: toward a multilevel theory of the legitimacy process. Acad Manag Rev 2015 Jan;40(1):49-75. [doi: 10.5465/amr.2013.0318]
129. Haack P, Sieweke J. The legitimacy of inequality: integrating the perspectives of system justification and social judgment. J Manag Stud 2017 Nov 29;55(3):486-516. [doi: 10.1111/joms.12323]
130. Johnson C, Dowd TJ, Ridgeway CL. Legitimacy as a social process. Annu Rev Sociol 2006 Aug 01;32(1):53-78. [doi: 10.1146/annurev.soc.32.061604.123101]
131. Haraway D. When species meet: staying with the trouble. Environ Plan D 2010 Jan 01;28(1):53-55. [doi: 10.1068/d2706wsh]
132. Bourdieu P, Nice R. Outline of a Theory of Practice. Cambridge, UK: Cambridge University Press; 1977.
133. Star SL, Ruhleder K. Steps toward an ecology of infrastructure: design and access for large information spaces. Inf Syst Res 1996 Mar;7(1):111-134. [doi: 10.1287/isre.7.1.111]
134. Star SL. The ethnography of infrastructure. Am Behav Sci 1999 Nov 01;43(3):377-391. [doi: 10.1177/00027649921955326]
135. Monat J, Gannon T. What is systems thinking? A review of selected literature plus recommendations. Am J Syst Sci 2015 Jul;4(1):11-26. [doi: 10.5923/j.ajss.20150401.02]
136. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Inf Syst J 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]
137. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 01;27(3):491-497 [FREE Full text] [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
138. Nadal C, Sas C, Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. J Med Internet Res 2020 Jul 06;22(7):e17256 [FREE Full text] [doi: 10.2196/17256] [Medline: 32628122]
139. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar 19;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
140. Davis FD. A technology acceptance model for empirically testing new end-user information systems : theory and results. Massachusetts Institute of Technology. 1986. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2023-10-02]
141. Manski CF. The use of intentions data to predict behavior: a best-case analysis. J Am Stat Assoc 1990 Dec;85(412):934-940. [doi: 10.2307/2289590]

142. Conner M, Norman P. Understanding the intention-behavior gap: the role of intention strength. Front Psychol 2022 Aug 4;13:923464 [FREE Full text] [doi: 10.3389/fpsyg.2022.923464] [Medline: 35992469]
143. Peiris D, Praveen D, Johnson C, Mogulluru K. Use of mHealth systems and tools for non-communicable diseases in low- and middle-income countries: a systematic review. J Cardiovasc Transl Res 2014 Nov 11;7(8):677-691. [doi: 10.1007/s12265-014-9581-5] [Medline: 25209729]
144. Wehrens R, Bal R. Health programs struggling with complexity: a case study of the Dutch 'PreCare' project. Soc Sci Med 2012 Jul;75(2):274-282. [doi: 10.1016/j.socscimed.2012.02.052] [Medline: 22571889]
145. Verran H. Number as an inventive frontier in knowing and working Australia's water resources. Anthropol Theory 2010 Jun 07;10(1-2):171-178. [doi: 10.1177/1463499610365383]
146. Black J, Murray AD. Regulating AI and machine learning: setting the regulatory agenda. Eur J Law Technol 2019;10(3).
147. Pace J, Pearson SA, Lipworth W. Improving the legitimacy of medicines funding decisions: a critical literature review. Ther Innov Regul Sci 2015 May 30;49(3):364-368. [doi: 10.1177/2168479015579519] [Medline: 30222405]
148. Brown B, Cheraghi-Sohi S, Jaki T, Su TL, Buchan I, Sperrin M. Understanding clinical prediction models as 'innovations': a mixed methods study in UK family practice. BMC Med Inform Decis Mak 2016 Aug 09;16:106 [FREE Full text] [doi: 10.1186/s12911-016-0343-y] [Medline: 27506547]
149. Benniche S. Using blockchain technology to recycle cancer drugs. Lancet Oncol 2019 Jun;20(6):e300. [doi: 10.1016/s1470-2045(19)30291-8]
150. Mandeville KL, Barker R, Packham A, Sowerby C, Yarrow K, Patrick H. Financial interests of patient organisations contributing to technology assessment at England's National Institute for Health and Care Excellence: policy review. BMJ 2019 Jan 16;364:k5300 [FREE Full text] [doi: 10.1136/bmj.k5300] [Medline: 30651227]
151. Pocock SJ, Stone GW. The primary outcome is positive - is that good enough? N Engl J Med 2016 Sep 08;375(10):971-979 [FREE Full text] [doi: 10.1056/NEJMra1601511] [Medline: 27602669]
152. Myers SP, Xue CC, Cohen MM, Phelps KL, Lewith GT. The legitimacy of academic complementary medicine. Med J Aust 2012 Jul 16;197(2):69-70. [doi: 10.5694/mja12.10491] [Medline: 22794038]
153. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251. [doi: 10.1016/s1470-2045(20)30021-8]

Abbreviations

CIS: critical interpretive synthesis
EBM: evidence-based medicine
IVF: in vitro fertilization
MAS: medical anthropology and sociology
OMS: organization and management studies
STS: science and technology studies
TAM: Technology Acceptance Model
WoS: Web of Science

Edited by A Kushniruk; submitted 19.05.23; peer-reviewed by M Barbosa, A Geampana, G Berntsen; comments to author 11.12.23; revised version received 24.01.24; accepted 01.12.24; published 05.03.25.

Please cite as:
Howe S, Uyl-de Groot C, Wehrens R
Legitimacy as Social Infrastructure: A Critical Interpretive Synthesis of the Literature on Legitimacy in Health and Technology
JMIR Hum Factors 2025;12:e48955
URL: https://humanfactors.jmir.org/2025/1/e48955
doi:10.2196/48955
PMID:40053717

Original Paper

Benefits and Barriers to mHealth in Hypertension Care: Qualitative Study With German Health Care Professionals

Susann May[1,2], MPH; Felix Muehlensiepen[1,2], Dr rer med; Eileen Wengemuth[1,2], Dr rer nat; Frances Seifert[1,2], MSc; Martin Heinze[3], Prof Dr Med; Dunja Bruch[2], MSc; Sebastian Spethmann[4,5,6], Prof Dr Med

[1]Brandenburg Medical School Theodor Fontane, Center for Health Services Research, Faculty of Health Sciences, Rüdersdorf, Germany

[2]Brandenburg Medical School Theodor Fontane, Department of Cardiovascular Surgery, Heart Center Brandenburg, Faculty of Health Sciences, Bernau, Germany

[3]Brandenburg Medical School, Immanuel Hospital Rüdersdorf, University Clinic for Psychiatry and Psychotherapy, Rüdersdorf, Germany

[4]Deutsches Herzzentrum der Charité, Department of Cardiology, Angiology and Intensive Care Medicine, Berlin, Germany

[5]corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Charité – Universitätsmedizin Berlin, Berlin, Germany

[6]German Center for Cardiovascular Research (DZHK), Partner Site Berlin, Berlin, Germany

Corresponding Author:
Susann May, MPH
Brandenburg Medical School Theodor Fontane
Center for Health Services Research, Faculty of Health Sciences
Seebad 82/83
Rüdersdorf, 15562
Germany
Phone: 49 3391 39 145 91
Email: susann.may@mhb-fontane.de

Abstract

Background: Digital health technologies, particularly mobile health (mHealth) apps and wearable devices, have emerged as crucial assets in the battle against hypertension. By enabling lifestyle modifications, facilitating home blood pressure monitoring, and promoting treatment adherence, these technologies have significantly enhanced hypertension treatment.

Objective: This study aims to explore the perspectives of health care professionals (HCPs) regarding the perceived benefits and barriers associated with the integration of mHealth apps into routine hypertension care. Additionally, strategies for overcoming these barriers will be identified.

Methods: Through qualitative analysis via semistructured interviews, general practitioners (n=10), cardiologists (n=14), and nurses (n=3) were purposefully selected between October 2022 and March 2023. Verbatim transcripts were analyzed using qualitative content analysis.

Results: The results unveiled 3 overarching themes highlighting the benefits of mHealth apps in hypertension care from the perspective of HCPs. First, these technologies possess the potential to enhance patient safety by facilitating continuous monitoring and early detection of abnormalities. Second, they can empower patients, fostering autonomy in managing their health conditions, thereby promoting active participation in their care. Lastly, mHealth apps may provide valuable support to medical care by offering real-time data that aids in decision-making and treatment adjustments. Despite these benefits, the study identified several barriers hindering the seamless integration of mHealth apps into hypertension care. Challenges predominantly revolved around data management, communication contexts, daily routines, and system handling. HCPs underscored the necessity for structural and procedural modifications in their daily practices to effectively address these challenges.

Conclusions: In conclusion, the effective usage of digital tools such as mHealth apps necessitates overcoming various obstacles. This entails meeting the information needs of both HCPs and patients, tackling interoperability issues to ensure seamless data exchange between different systems, clarifying uncertainties surrounding reimbursement policies, and establishing the specific clinical benefits of these technologies. Active engagement of users throughout the design and implementation phases is crucial for ensuring the usability and acceptance of mHealth apps. Moreover, enhancing knowledge accessibility through the provision of easily understandable information about mHealth apps is essential for eliminating barriers and fostering their widespread adoption in hypertension care.

Trial Registration: German Clinical Trials Register DRKS00029761; https://drks.de/search/de/trial/DRKS00029761

International Registered Report Identifier (IRRID): RR2-10.3389/fcvm.2022.1089968

(JMIR Hum Factors 2025;12:e52544) doi:10.2196/52544

KEYWORDS

hypertension; mHealth apps; digital health; physicians; nurses; HCP; qualitative interviews; health care professional; cardiologists; mHealth; Germany; general practitioners; blood pressure monitoring; qualitative study; qualitative content analysis

Introduction

Hypertension is a significant health problem affecting millions of people worldwide [1]. In Germany, too, hypertension is one of the most common diseases affecting a large proportion of the population. Prevention and effective treatment of hypertension are therefore crucial to prevent associated serious complications such as ischemic heart disease, strokes, and renal disease, and to improve the quality of life of those affected [2-4].

In recent years, the development of digital health technologies has experienced a considerable upswing, opening up promising perspectives for the prevention and management of hypertension [5]. Mobile health (mHealth) apps, wearables, and other digital tools enable patients to monitor their blood pressure levels, track lifestyle changes, and set health goals. These technological advances also offer health care professionals (HCPs) new opportunities to improve patient care and promote healthier lifestyles [6]. mHealth apps in particular have been shown to improve the treatment of hypertension [7].

In Germany, the following professional groups are primarily involved in the treatment of people with hypertension. General practitioners play a central role in the diagnosis, treatment, and long-term care of patients with hypertension. They are often the first point of contact for patients with hypertension and coordinate further care. Cardiologists specialize in the diagnosis and treatment of heart disease, including hypertension. They can treat complex cases of hypertension and offer specialized diagnosis and treatment options. Internists can also play a role in the care of patients with hypertension, especially if there are comorbidities or complex medical problems. Nursing professionals (nurses, nursing assistants) play a crucial role in the care of patients with hypertension. They help implement blood pressure control measures, educate patients, and help monitor symptoms and side effects [8]. In October 2020, physicians in Germany were allowed to prescribe digital therapeutics (DiGA [Digitale Gesundheitsanwendungen]) in standard care for the first time [9]. DiGAs can be prescribed by physicians for a 3-month use period and is fully reimbursed by the statutory insurance companies. To obtain approval, mHealth providers must demonstrate the safety, functionality, quality, data security, and a fundamental benefit in a clinical study to obtain DiGA status. Patients receive a written prescription for a DiGA based on respective indications (eg, depression) and have to send it to their insurance company.. At the time this study was conducted, there was no DiGA that is specific for the prevention of hypertension. In perspective, however, it is expected that mHealth apps for the prevention of hypertension will obtain DiGA status and will be made available to patients. In their recent study, Dahlhausen et al [10] pointed out the significant role of HCPs, specifically physicians, in promoting mHealth use among patients. In order to understand the implementation of mHealth into hypertension care, it is necessary to understand the perspectives of physicians, nurses, and medical assistants and their perception of benefits and challenges, as well as their information needs, in order to best integrate mHealth into hypertension care.

Nevertheless, a certain frustration is noticeable among physicians in Germany. Not least because there have been repeated technical problems in the past with the connection to the German telematics infrastructure (TI). The TI comprises a range of technical components that are relevant for operation in medical practices. These include the practice management system (Praxis Verwaltungssystem), which supports the organization and documentation of practice tasks, and the connector, a piece of hardware that enables access to the TI and communicates with the eHealth card terminal and the Praxis Verwaltungssystem via a secure network. The eHealth card terminal is used to use the electronic health card and to check practice ID cards. Other components include mobile card terminals, practice or institution ID cards, a virtual private network access service, and the electronic health professional card, which is required for various applications such as the electronic doctor's letter and prescriptions [11]. However, half of the physicians have to struggle with technical problems at least once a week, compared to 36% in 2020. Overall, frustration with the digitization process has increased [12].

While there is a wealth of research focusing on the patient perspective regarding digital health tools [13-15], there remains a lack of comprehensive evidence on how physicians perceive and integrate these technologies into their clinical practice. Despite the promising potential of digital health tools, the potential obstacles and barriers to widespread adoption of these technologies in the clinical setting have not been thoroughly explored. A deeper understanding of the structural and individual factors that influence the implementation and adoption of digital prevention approaches is imperative. Integrating digital prevention approaches into the health care system requires a clear delineation of the role and positioning of these technologies within the broader health care landscape. A thorough investigation of how digital health tools can be effectively integrated into the existing health care infrastructure and how they are perceived by the various stakeholders in the health care sector is crucial.

Addressing these research gaps can contribute significantly to a more nuanced understanding of how digital health and specifically mHealth apps can be optimally used for hypertension prevention. It can also identify potential barriers and pave the way for strategies to improve clinical adoption and seamless integration of these technologies. In this study, we explore the following research questions: Which benefits

do HCPs see in mHealth apps for prevention? What are possible structural and individual barriers to the use of mHealth apps? How can the challenges be met?

Methods

Study Design

This paper reports on our exploration of the HCPs' perspectives regarding mHealth apps in hypertension prevention. We conducted in-depth interviews with general practitioners, cardiologists, and nurses. This study is part of the DiPaH project [16], which examines structural and individual factors in different stakeholders that influence the use of digital preventive measures in patients with arterial hypertension in Germany.

Ethical Considerations

This study was approved by the Ethics Committee of the Brandenburg Medical School Theodor Fontane (E-02-20220620). All experimental protocols were approved by a named institutional or licensing committee. All methods were carried out in accordance with relevant guidelines and regulations. Participants were informed verbally and in writing about the purpose, the procedure, the significance of the study, and the benefits and risks that may be associated with it and had the opportunity to ask questions. They were also informed that they had the right to withdraw their consent to participate in the study at any time, either verbally or in writing, without giving reasons. They were also informed that personal data would be recorded and stored, whereby the data would be pseudonymized, but that no data would be published that could be used to identify the person. For security reasons, the data received from the participants were always stored in a password-protected folder on a secure desktop computer. Written consent was obtained after participants were given the opportunity to ask questions. Patients were not involved in designing this study. Participants were offered €75.00 (approximately US $77.89) as an incentive for their participation in the study.

Recruitment

Participants were selected using purposive sampling [17]. We included HCPs who were currently actively involved in the treatment of people with hypertension, such as general practitioners, cardiologists, and nurses. The participants were selected according to defined categories, aiming to generate a heterogeneous sample with the broadest variety of different perspectives [18]. The following categories were included: region of practice (rural vs urban), main area of practice (outpatient vs inpatient), and gender (male vs female). A further inclusion criterion was interest and willingness in participating in an interview. Participants were recruited both by systematic invitation and by snowball sampling. The following sources were used for recruitment (Textbox 1).

Textbox 1. Sources used for recruitment.

- Social media accounts of the university (Brandenburg Medical School Theodor Fontane), such as Facebook, Instagram, and Twitter
- Association of Statutory Health Insurance Physicians Brandenburg (KVBB): newsletter
- General practitioners' association Brandenburg: information study participation at events
- Personal contacts among colleagues (snowball sampling)
- Interested persons contacted the study team by telephone or e-mail, and it was checked in advance whether they could be included before an interview was conducted

Data Collection

A preliminary semi-structured interview guide was drafted by a multiprofessional team (SM, FM, DB, and SSp). The semistructured interview guide consisted of open-ended questions that explored how the participants are attuned to digital applications and mHealth apps, what benefits and barriers they see in digital technologies for prevention, what information needs they have, and how these can be met. The focus was on hypertension apps with the following functions: educating about hypertension, monitoring blood pressure, and promoting adherence, although no specific apps were offered or recommended to participants. Sample interview questions included: Which mHealth apps in the context of hypertension treatment are used in your practice, or what experiences have you had with it? What benefits do you see in the use of mHealth apps? What barriers do you see in the use of mHealth apps? How well informed are you about mHealth apps, and where are knowledge deficits? In addition, sociodemographic data were collected, including profession, gender, age, number of inhabitants of the place of practice, duration of professional activity, and setting. To ensure clarity and relevance of the questions, we did a pilot test of the interview guide with 5 eligible participants recruited from clinics and outpatient physicians in the study's catchment area (refer to Multimedia Appendix 1).

All interviews were conducted via telephone during October 2022 to March 2023. The interviews were recorded and transcribed verbatim in accordance with data protection guidelines. Data collection continued until no substantially new findings emerged and saturation of content was reached. Saturation of content is defined as code saturation, when no additional issues are identified, and meaning saturation, when no further dimensions, nuances, or insights of issues can be found [19]. We chose to conduct 27 interviews, as this number has proven to be adequate in previous studies to reach theoretical saturation in the exploration of digital approaches [20,21]. Field notes (a short summary of the interview) were taken following each interview to ensure better comprehensibility of the interview situation. The field notes themselves were not included in the analysis.

Data Analysis

Qualitative analysis of the interviews was performed iteratively by the study team (SM and FM) based on Kuckartz's structured qualitative content analysis [22] using MAXQDA Analytics Pro 2022, Release (version 22.1.0; Verbi GmbH). After transcription of the audio material, the analysis began with a screening of the interview texts, whereupon the interviews were coded. Relevant text passages from the interview material were coded according to a deductive-inductive procedure. Main categories and subcategories were formed inductively from the codes. Until a common understanding of all the emerging categories was achieved, consensus discussions were held continuously in the research group. Other categories were developed deductively based on the research questions and merged into the coding tree. By then, the process of gathering data had already concluded. Two researchers (SM and FM) separately applied the established category system to analyze the complete data set, ensuring that the process could be traced and replicated. The data was analyzed in German. To present the findings, significant excerpts from the transcriptions were chosen as representative quotes. These quotes were translated into English by native speakers and incorporated into the manuscript. The manuscript has been compiled in accordance with the COREQ (Consolidated Criteria for Reporting Qualitative Research; refer to Multimedia Appendix 2) [23].

Results

In total, 27 interviews were conducted and analyzed until theoretical saturation was reached. The mean duration of the interviews was 42 (range 33-56) minutes. The mean age of the participants (N=27) was 50 (range 35-74) years. Most participants were female (14 female/13 male). A total of 14 cardiologists, 3 nurses, and 10 general practitioners participated. In total, 9 persons worked in the inpatient setting and 17 persons worked in the outpatient setting. One person worked in both settings. Four of the HCPs had already had experience with digital applications in daily practice or were also working with them. Detailed characteristics of study participants are shown in Table 1.

Table 1. Detailed characteristics of study participants.

Characteristics	Participants (N=27), n (%)
Professional group	
General practitioner	10 (37)
Cardiologist	14 (52)
Nurse	3 (11)
Duration of professional activity	
>10	9 (33)
10-20	8 (30)
21-30	7 (26)
31-40	3 (11)
Gender	
Female	14 (52)
Male	13 (48)
Age (years)	
>40	2 (7)
40-50	11 (41)
51-60	11 (41)
<60	3 (11)
Number of inhabitants	
>10,000	2 (7)
10-100,000	11 (41)
100,001-1 million	4 (15)
<1 million	10 (37)
Setting[a]	
Inpatient	10 (37)
Outpatient	18 (67)

[a]One participant works in both the inpatient and outpatient settings.

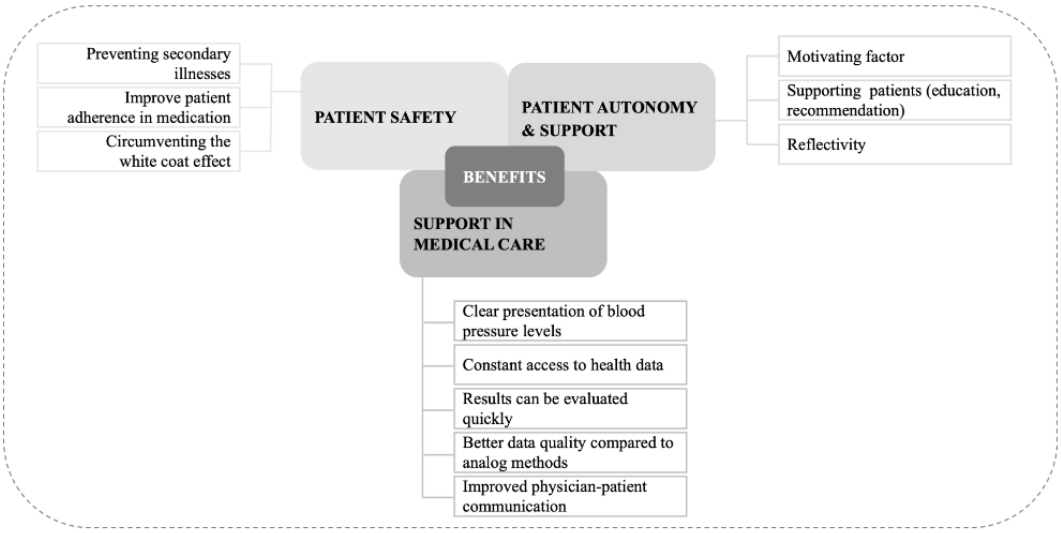
Themes

The analysis explored 7 key themes, categorized into benefits and barriers. Themes identified as benefits included (1) patient safety, (2) patient autonomy and support, and (3) support in medical care. Themes identified as barriers were organized by the following context: (4) handling of data, (5) in the context of communication, (6) in daily routines, and (7) in dealing with the systems.

Benefits

Although few of the HCPs use or have used mHealth apps in their daily practice, they were able to mention benefits they expected to gain from using mHealth apps. In the context of benefits, 3 overarching themes could be identified: potential to increase patient safety, patient autonomy, and support in medical care (Figure 1 and Multimedia Appendix 3).

Figure 1. Coding tree of the benefits regarding mHealth apps in hypertension prevention. mHealth: mobile health.



A different number of subcategories could be assigned to the 3 main categories. The following 3 subcategories were assigned to the “Patient safety” category: preventing secondary illnesses, improving patient adherence in medication, and circumventing the white coat effect. The following 3 subcategories were assigned to the category “Patient autonomy and support”: motivating factor, supporting patients, and reflectivity. The following 5 subcategories were categorized under the category “Support in medical care”: clear presentation of blood pressure levels, constant access to health data. Results can be evaluated quickly, better data quality compared to analog methods, and improved physician-patient communication.

Patient Safety

Respondents see the potential of mHealth apps primarily in the prevention of secondary diseases. In addition, participants see the potential of mHealth apps to improve patient adherence to treatment in terms of medication adherence, which in turn has an impact on patient safety.

The respondents attribute a better representation to the blood pressure levels measured at home and subsequently documented. The interviewees hope that through the use of mHealth apps, patients will consistently document their blood pressure and thus always have an overview of their blood pressure levels. This is because, in contrast to the blood pressure levels measured in the practice, they seem to correspond more closely to the reality of life. The “white coat effect” is thus circumvented.

Patient Autonomy and Support

An mHealth app can offer a variety of features that serve as a motivating factor for patients from the perspective of health care providers. Gamified elements, progress tracking, and reward systems can provide incentives for developing and maintaining health-promoting behaviors, strengthening patients' self-motivation, and fostering engagement in their health.

From the HCP’s perspective, an mHealth app can be a valuable addition to medical care by providing support to patients in several ways. First, the mHealth app can provide personalized advice based on the patient's individual needs and health goals. This can include dietary recommendations, lifestyle changes, or stress management tips. Second, mHealth apps can act as reminders, reminding patients to take their medication, which could improve adherence to therapy.

An mHealth app can also include reflectivity and help patients comprehend information about their blood pressure to their behaviors, which can support them reduce or avoid unfavorable behaviors.

Support in Medical Care

By using mHealth apps, the HCPs assume that blood pressure levels can be presented in a clear and understandable way. mHealth apps allow patients to access their levels on their smartphones or other digital devices. Patients can track the development of their blood pressure over time and identify possible trends or abnormalities so that they can seek medical help timely if necessary.

Patients have constant access to their current health data. This enables real-time tracking of their health status, even if they are not undergoing immediate medical treatment.

mHealth apps provide quick results that can be evaluated immediately. Physicians can thus make timely decisions about the treatment and care of their patients.

HCPs attribute better data quality to mHealth apps compared to analog methods.

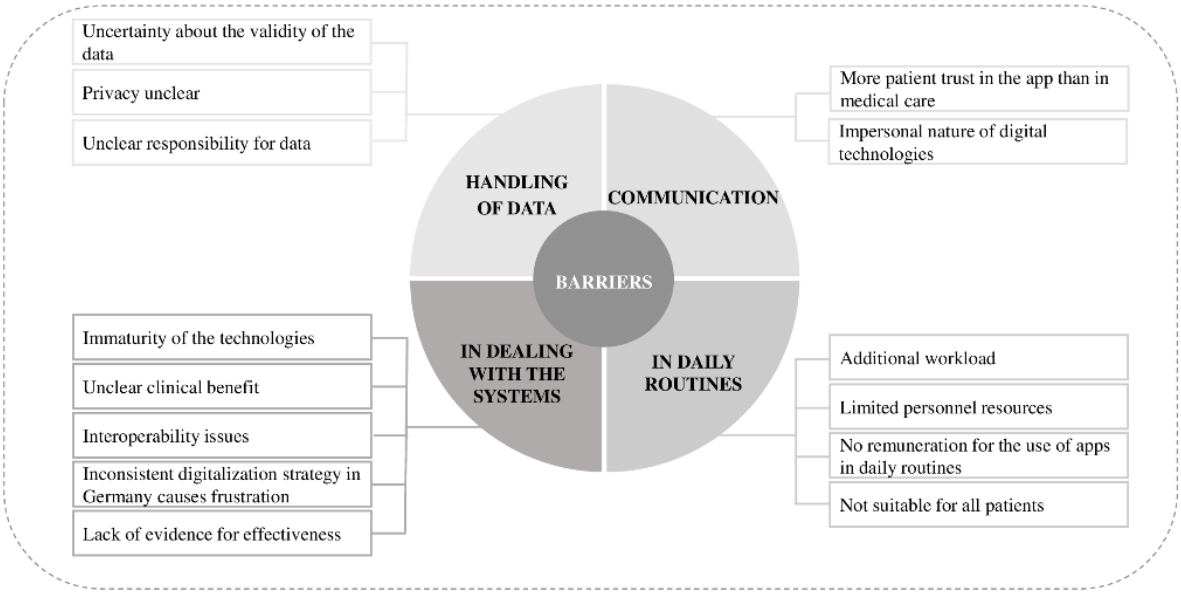
Using mHealth apps, patients can send the data they collect in their daily environment directly to the medical facility before visiting the physician. This allows the physician to check the blood pressure levels in advance and prepare specifically for the conversation with the patient. The time saved helps to make the physician-patient conversation more efficient and targeted so that relevant questions and concerns of the patient can be better addressed.

The benefits initially included constant access to health data, which was particularly mentioned by participants in the outpatient sector. This group also appreciated the motivating factor that the apps could provide to support them in their health routines. Participants in large cities also preferred the app's clear display of their blood pressure readings, which helped them to better understand their health data. The time saved in doctor-patient communication was also mentioned as a benefit by the outpatient participants.

Barriers

We identified barriers in 4 different aspects of the HCPs work life: handling of data, in the context of communication, in daily routines, and in dealing with the systems (Figure 2 and Multimedia Appendix 4). A different number of subcategories could be assigned to the 4 main categories. The following 3 subcategories were assigned to the category “Handling of data”: uncertainty about the validity of the data, privacy unclear, and unclear responsibility for data. The following two subcategories were assigned to the category “in the context of communication”: more patient trust in the app than in medical care and the impersonal nature of digital approaches. The following 4 categories were assigned to the category “in daily routines”: additional workload, limited personnel resources, no remuneration for the use of apps in daily routines, and not suitable for all patients. The following 5 subcategories were assigned to the category “in dealing with the systems”: immaturity of the technologies, unclear clinical benefit, interoperability issues, and inconsistent digitalization strategy in Germany, which causes frustration and lack of evidence for effectiveness.

Figure 2. Coding tree of the barriers regarding mHealth apps in hypertension prevention. mHealth: mobile health.



Handling of Data

One of the main barriers to the use of apps for blood pressure monitoring is uncertainty about the validity of the data collected. Concerns about the accuracy of data documented by patients can undermine HCPs’ confidence in the data and willingness to use such tools.

Another important concern related to the use of mHealth apps in the treatment of hypertension is privacy. Fear of data breaches or unauthorized access to patient information may limit the use of such tools.

Unclear responsibility for data presents another hurdle. Physicians and nurses feel uncertain about who is responsible for managing and interpreting the data they collect. The wealth of data generated by mHealth apps can be overwhelming, and the lack of clear structures and policies for appropriate use of the data can be challenging.

In the Context of Communication

The HCPs expect that a potential barrier to communication when using mHealth apps is the tendency of some patients to trust the mHealth app more than their medical care. This misconception could lead to patients relying solely on the mHealth app for solutions or seeing its advice as better than that of their physician. This could hinder communication with HCPs and lead to a lack of collaboration between patients and physicians.

Another barrier perceived by HCPs is the impersonal nature of mHealth apps. In contrast to direct interaction, the use of mHealth apps could lead to a lower intensity of exchange between patients and HCPs. In addition, mHealth apps may be perceived as less empathetic and less tailored to the individual needs of patients.

In Daily Routines

The HCPs apprehend that introducing mHealth apps in medical practice can initially lead to an additional workload for them. The introduction of new technologies requires training, the adaptation of workflows, and integration into existing IT systems, which can initially lead to time and resource constraints.

In many health care settings, resources are lacking in order to engage HCPs in the active use of mHealth apps in practice. Adoption and integration of new technologies requires time, training, and technical support, which can be challenging in an already strained health care system. Limited capacity impedes readiness to use mHealth apps for hypertension care.

At present, the use of mHealth apps is not sufficiently reimbursed in everyday health care in Germany. HCPs do not receive additional compensation or incentives for the use and deployment of digital tools in the care of hypertensive patients. This lack of compensation may cause some HCPs to be hesitant to use digital tools, as it may entail additional tasks without financial compensation.

Although mHealth apps offer many benefits, they may not be suitable for all patients. Older people or those with limited digital capabilities may have difficulty using digital applications. In addition, some patients, especially those with hypertension, may overmeasure their blood pressure, which can negatively impact their blood pressure levels. According to HCPs, digital tools for monitoring blood pressure may be less appropriate for such patients.

In Dealing With the Systems

A major barrier to the use of mHealth apps in hypertension care is the immaturity of the technologies. HCPs suspect that mHealth apps do not work reliably, have bugs, or do not have all the necessary features to meet the demands of medical practice. As a result, they may be reluctant to recommend digital technologies for fear that the systems will not exist at a later date.

Another barrier to the use of mHealth apps in hypertension care is the unclear actual utility and clinical effectiveness of such tools from the perspective of HCPs.

The diversity of mHealth apps and their interfaces is another barrier for HCPs. The incompatibility or lack of interoperability between systems can hinder smooth data exchange and integration.

The insufficient as well as inconsistent digitalization strategy in Germany is a source of frustration for HCPs. There is often a lack of a clear, long-term vision and of a structured approach to drive the integration of digital technologies in health care. The development and implementation of digital health solutions is often fragmented and inconsistent, leading to confusion, ineffective solutions, and a return to analog methods. These negative experiences are a significant barrier to the adoption of digital tools.

The actual effectiveness and long-term impact of mHealth apps in treating hypertension are poorly researched or demonstrated from the perspective of HCPs. The lack of evidence on the clinical effectiveness of such tools raises doubts and influences their willingness to use them.

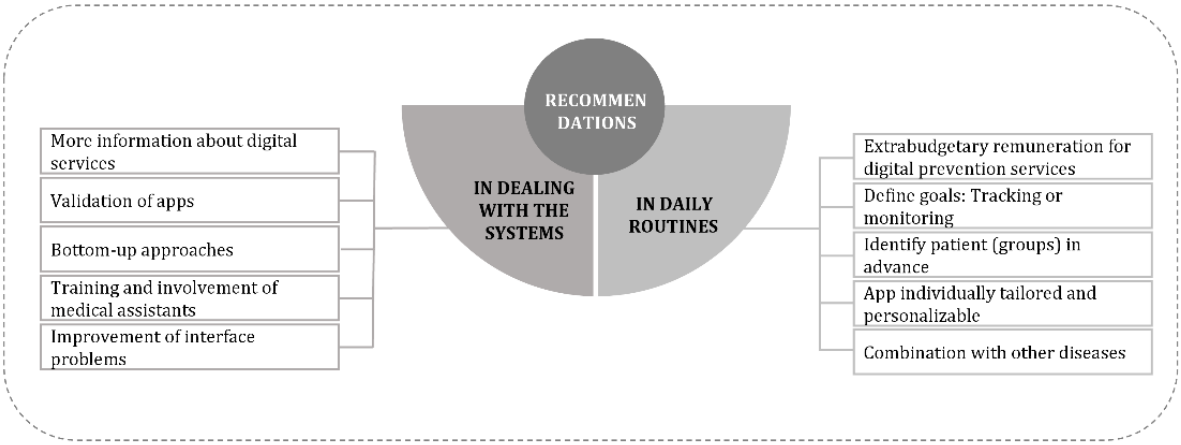
Specific Aspects in Various Subgroups

Participants from the outpatient sector in particular pointed out interoperability problems that could hinder the smooth exchange of data. In addition, the unclear clinical benefits of the apps were not recognized by everyone, especially not by the participants from the outpatient sector. It was also noted from the outpatient sector that not all apps are suitable for every patient. Concerns about data protection were also expressed by participants in the outpatient sector, who were worried about the security of their personal health data. Finally, the lack of reimbursement for the use of health apps from the outpatient sector was cited as a further barrier. Women expressed concerns about the validity of the data collected by the apps.

Recommendations for Successful Integration of mHealth Apps Into the Health Care Landscape From HCPs' Perspective

In the context of recommendations, 2 themes could be identified: in daily routines and in dealing with the systems (Figure 3 and Multimedia Appendix 5). A different number of subcategories could be assigned to the 2 main categories. The following 5 categories were assigned to the category "in daily routines": extrabudgetary remuneration for digital prevention services, defining goals: tracking or monitoring, identifying patient (groups) in advance, app individually tailored and personalizable, and combination with other diseases. The following 5 subcategories were assigned to the category "in dealing with the systems": more information about digital services, validation of apps, bottom-up approaches, training and involvement of medical assistants, and improvement of interface problems.

Figure 3. Coding tree of the recommendations for successful integration of mHealth apps into the health care landscape. mHealth: mobile health.



In Daily Routines

One of the fundamental challenges in integrating hypertension mHealth apps into routine medical practice is that the time spent using and counseling patients is not adequately compensated. To address this problem and increase HCPs' motivation to use digital prevention services, extrabudgetary reimbursement for the implementation and support of these apps in HCPs' practices should be considered. According to the participants, such remuneration would ensure that HCPs are appropriately compensated and thus create an incentive to increasingly integrate mHealth apps into their daily work.

It is crucial to define clear goals for the use of mHealth apps in hypertension prevention. HCPs should work with patients to determine which parameters or data should be collected by the mHealth app and which specific health goals are being pursued. The HCPs assume that a distinction can be made between tracking and monitoring approaches. While tracking helps patients to collect data on their blood pressure, lifestyle, and other relevant parameters themselves, monitoring enables the physician to regularly monitor and analyze the collected data. The clear definition of goals facilitates the use of the mHealth apps and ensures that the collected data can be used meaningfully and effectively.

The HCPs mentioned that not all patients benefit equally from mHealth apps. It is important to identify in advance which patients or patient groups could benefit most from a hypertension prevention mHealth app. Some patients may already be well informed and motivated to use digital health technologies on their own, while others may need more support and guidance. By identifying those patients who could best benefit from mHealth app use, clinicians can target their mHealth services more effectively and efficiently.

Another important improvement is that hypertension apps should be customized and personalizable. Each patient is unique, and the mHealth app should therefore be customizable to each individual's specific needs, health goals, and preferences. Personalization options could include selection of relevant health goals, medication reminders, or customization of the app interface to the user's personal preferences. By tailoring the app to each patient's individual needs, the user experience is likely to be enhanced and the likelihood of long-term use increased.

Hypertension often occurs in combination with other conditions, such as diabetes or obesity. To improve the care of multimorbid patients, mHealth apps for hypertension could be combined with other health applications to provide comprehensive and integrated care.

In Dealing With the Systems

It is important that HCPs have access to comprehensive and reliable information about the functionalities, efficacy, safety, and privacy policies of the various hypertension apps. A transparent and evidence-based presentation of this information can help to address concerns about the quality and reliability of the apps and stimulate physicians' interest in using these digital tools.

Careful validation of mHealth apps is essential to ensure that hypertension management apps actually deliver the promised benefits and provide medically robust data. HCPs need reliable information about the scientific evidence and clinical validity of the apps in order to effectively integrate them into their clinical practice.

To improve the adoption and integration of apps for hypertension, it is crucial to promote bottom-up approaches. This means including the user perspective, including physicians and medical staff, in the development and implementation of mHealth apps from the very beginning. Involving users in the development process makes it possible to adapt mHealth apps to the specific needs and requirements of physicians and ensure high usability.

Successful integration of hypertension apps requires not only training of physicians but also the involvement of other HCPs, especially medical assistants. Medical assistants play an important role in helping physicians use mHealth apps and interact with patients. Extensive training of medical assistants on how the mHealth apps work and are useful, as well as how to effectively use the data collected, can help ensure that the integration of the apps is seamless in the practice workflow.

Interoperability and compatibility between different mHealth apps and existing practice IT systems are critical. Often,

physicians and medical assistants face the challenge that the interfaces between the mHealth apps and existing IT systems are not sufficiently optimized, which can lead to technical difficulties and data inconsistencies. Targeted improvement of the interface issues enables physicians to work smoothly with the mHealth apps and seamlessly integrate the captured data into their practice workflows.

Discussion

Principal Results

This qualitative study is the first to explore the factors that enable or hinder the use of mHealth apps for the prevention of hypertension from the perspective of general practitioners, cardiologists, and nurses in Germany. Until now, mHealth apps for the prevention of hypertension are only fragmentarily used in the care of hypertensive patients. From the perspective of the caregivers, the main potentials of mHealth apps lie in increasing patient safety, patient autonomy, and support in medical care. However, the presented results point to major challenges in data handling, patient-HCP communication, integration into standard care, and interoperability. Measures such as extra-budgetary remuneration for digital preventive services, setting targets through tracking or monitoring, identifying patient groups in advance, adapting and personalizing mHealth apps, and combining them with other diseases could, in the view of the HCPs, help to overcome the barriers. In dealing with the systems, the HCPs mention aspects such as more information about digital services, validation of mHealth apps, bottom-up approaches, training and involvement of medical assistants, and improvement of interface problems.

Comparison With Prior Work

Our results are in line with earlier research in mHealth and beyond: physicians and nurses recognize the added values and positive potential of mHealth apps in supporting patients, fostering patient autonomy, and improving medication adherence [24,25]. The attributed benefits are supported by evidence: for instance, Li et al [26] report improvements in self-management behavior and medication adherence in adults using mHealth. Furthermore, self-monitoring of hypertension-related behaviors via smartphone apps combined with tailored advice has a modest but potentially clinically significant effect on blood pressure reduction [27].

Despite the advantages of mHealth, as of yet the barriers hinder their integration into the medical routine. One of the main concerns is the handling of data. Physicians frequently doubt the validity of data collected by patients and have concerns regarding data privacy and security. This can impede trust in the data and willingness to use such mHealth apps.

The mode of communication between physicians and patients can also be influenced by the use of mHealth apps. Some patients may tend to place more trust in the instructions provided by the mHealth app than in the recommendations of their medical caregivers. This could hinder communication with HCPs and impede collaboration between patients and physicians. However, studies have already demonstrated that there are contrary findings to the presumed concerns: findings of a systematic, narrative review show a generally positive influence of mHealth apps on physician-patient communication and relationships, which at times was correlated with better health outcomes [28]. And perhaps the use of apps can make visits more efficient, because blood pressure levels can be transmitted to the practice in advance, and conversations can be focused on essential aspects.

In the interviews, it is stated that mHealth consumes more resources and deteriorates patient communication. In rheumatology, digital health and especially mHealth are considered resource-saving. Digitization creates more time for patient interaction, which enhances the quality of care. If questions have already been answered beforehand, there may be extra time in the conversation with the patient to focus on essential aspects, which can positively impact the perceived quality of care [29]. Furthermore, it is emphasized in rheumatology care that this would allow more time for patient consultations. This is a contradiction. However, rheumatology is more digitized compared to hypertension care because of the greater shortage of personnel and disease burden.

Furthermore, apps can initially mean additional workload for HCPs. The introduction of new technologies requires training, adaptation of workflows, and integration into existing IT systems, which can lead to time and resource constraints. In addition, due to limited resources, many medical facilities may not be adequately prepared to actively integrate digital technologies into their practice. To be able to use digital technologies in their daily routine, physicians also need the appropriate digital skills [30]. In a recent measurement of physicians' professional digital health literacy, they scored 53.1 out of a possible 100 points [31]. For nurses, this figure is only slightly higher at 54.5 points. The most difficult task for physicians and nurses is “helping patients assess the trustworthiness of the digital health information they find,” followed by “helping patients find the digital health information that is relevant to them.” In the future, physicians and nurses will need support especially in this area so that the digital transformation in health care can be driven forward.

A lack of integration of digital tools into the reimbursement structure of the health care system in Germany is also a barrier to the use of such applications. Currently, physicians only receive remuneration for prescribing DiGAs, which does not translate to monitoring apps that have not obtained the DiGA status. There are ongoing discussions about which remuneration model is suitable for mHealth. Labinsky et al [32] reported that only 13% of patients that were prescribed a DiGA completed the whole program. This may also be due to a lack of incentives for physicians to ask if the apps are being used. Considering the low adherence with DiGA, performance-based compensation appears to be appropriate.

Implications

Several recommendations for the integration of mHealth apps into medical practice can be derived from the results. It is important to provide HCPs with low-threshold access to comprehensive and reliable information about available mHealth apps to address their concerns about quality and privacy. Validation of mHealth apps for effectiveness and clinical

validity is critical to convince physicians that these apps provide real value [33]. Bottom-up approaches or participatory study designs that include the user perspective, including physicians and medical assistants in the development process, can help ensure that apps are better tailored to the needs and requirements of medical practice and have a higher usability.

Involving HCPs, especially medical assistants, is also important as they play a key role in helping physicians use mHealth apps and interact with patients. Extensive training of medical staff on how the apps work and their benefits can help ensure that the integration of the apps is seamless in the practice workflow.

Our results underline the importance of staggered use of the mHealth app for effective blood pressure control in patients with hypertension. A high frequency of use at the beginning of treatment helps both patients and health care providers to quickly identify and implement individualized strategies to lower blood pressure. Once the target values have been stably achieved, the frequency of use can be gradually reduced so that the mHealth app can serve as a monitoring tool without unnecessarily increasing the time required by medical staff. This flexible handling could allow the mHealth app to be optimally integrated into everyday clinical practice while minimizing the burden on health care providers in terms of data management.

In addition, it is important to improve the interface issues between digital health apps and existing IT systems. Interoperability and compatibility between different digital apps and medical practice software systems are critical to ensure smooth integration and use of the apps. Targeted improvement of the interfaces can enable physicians to work smoothly with the apps and seamlessly integrate the collected data into their practice workflows.

The use of digital care services could be beneficial if the goal of use were addressed in advance in physician-patient communication. For example, use in the sense of self-tracking would not necessarily require medical intervention. Whereas the use in the sense of monitoring is associated with a control function, and physicians are actively involved in the process. Further studies are also needed to analyze the benefits and effectiveness of digital care services.

Based on the current findings and recommendations made by the HCPs, various stakeholders (eg, policy makers, health insurance companies, physicians' associations, associations of the scientific medical societies, patient organizations, health care researchers, or app developers) in different fields of the health care system should take action to push the integration of mHealth app in daily practice (Textbox 2).

Textbox 2. Recommendations for the integration of mobile health (mHealth) apps into medical practice.

Structural conditions

- Extrabudgetary remuneration for digital prevention services
- Improvement of interface problems
- Training and involvement of medical assistants
- More information about digital offers

App development

- Bottom-up approaches and user experience design
- App individually tailored and personalizable
- Combination with other diseases
- Validation of mHealth apps

Use in patient care

- Assessment of patients' eligibility to use mHealth
- Identify patient (groups) in advance
- Define goals: tracking or monitoring

Strength and Limitations

Our study is the first to investigate the benefits and barriers of using mHealth apps in preventing hypertension in Germany. We drew closely on the reality of care and the perspectives of the physicians and nurses. The qualitative interviews allowed for obtaining an in-depth understanding of the experiences with or without apps in preventing hypertension and expected advantages and disadvantages. However, there are certain limitations to our study. The results might be subject to a selection bias because perhaps people who were more interested in the topic took part in the study. Due to the sampling strategy, we may not have been able to reach everyone and did not record important aspects. A generalization of the results is therefore impossible. In addition, patients were not interviewed at this time.

The authors are aware that this general exploration can only be a first step given the breadth and diversity of the topic "hypertension mHealth apps." Further studies must be conducted in the context of specific health care situations to build upon these findings. These limitations directly translate into opportunities for further research: We are planning to extend

the study to a larger sample in Germany [16] and to validate the qualitative results with a questionnaire survey.

Further research should also focus on how to identify patients who may or may not be eligible for mHealth apps in hypertension care; how HCPs can be effectively informed about mHealth and digital health in general, as well as what reimbursement strategy will reinforce the effective use of mHealth in the care of chronic conditions.

Conclusions

HCPs recognize the benefits of mHealth apps for patient safety or to support disease management. However, at present, best practices for implementing mHealth apps in preventing hypertension do not exist. In order to be able to use digital tools such as mHealth apps efficiently, the barriers, such as information needs, interoperability issues, uncertain remuneration, and vague clinical effectiveness, need to be overcome. Above all, users need to be involved in design and implementation processes. Furthermore, low knowledge needs to be eliminated by offering access to low-threshold information on mHealth apps.

Acknowledgments

The authors would like to thank all interview partners and participants for their valuable contributions and time. We would like to express our special thanks to the Association of Statutory Health Insurance Physicians Brandenburg (KVBB) and the General Practitioners' Association Brandenburg for their support during the recruitment process. This work was supported by the Innovation Committee at the Federal Joint Committee (Innovation Fund, grant number 01VSF18014).

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). For further questions regarding the reuse of data, please contact the corresponding author (susann.may@mhb-fontane.de).

Authors' Contributions

SM made substantial contributions to the design of the work, the data acquisition and analysis, and the interpretation of the data; drafted the manuscript and had full access to all the data in the study; and takes responsibility for the integrity of the data and the accuracy of the data analysis. FM made substantial contributions to the design of the work, the data acquisition and analysis, and the interpretation of the data; drafted the manuscript and had full access to all the data in the study; and takes responsibility for the integrity of the data and the accuracy of the data analysis. EW made substantial contributions to the design of the work and the data acquisition and analysis and substantively revised the manuscript. FS made substantial contributions to the data analysis and substantively revised the manuscript. MH made substantial contributions to the data acquisition and analysis and substantively revised the manuscript. DB made substantial contributions to the design of the work, obtaining funding, and data acquisition and analysis, and substantively revised the manuscript. SSp made substantial contributions to the design of the work, obtaining funding, and data acquisition and analysis, and substantively revised the manuscript. All authors have approved the submitted manuscript. In addition, the authors have agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. SM and FM contributed equally to this work (shared first authorship). DB and SSp contributed equally to this work (shared last authorship).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Excerpt of the interview guide.
[DOCX File , 23 KB - humanfactors_v12i1e52544_app1.docx]

Multimedia Appendix 2
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[PDF File (Adobe PDF File), 56 KB - humanfactors_v12i1e52544_app2.pdf]

Multimedia Appendix 3
Coding tree with anchor quotes: benefits regarding mHealth apps in hypertension prevention. mHealth: mobile health.
[DOCX File , 15 KB - humanfactors_v12i1e52544_app3.docx]

Multimedia Appendix 4

Coding tree with anchor quotes: barriers regarding mHealth apps in hypertension prevention. mHealth: mobile health.
[DOCX File , 16 KB - humanfactors_v12i1e52544_app4.docx]

Multimedia Appendix 5
Coding tree with anchor quotes: recommendations for successful integration of mHealth apps into the health care landscape. mHealth: mobile health.
[DOCX File , 15 KB - humanfactors_v12i1e52544_app5.docx]

References

1. Forouzanfar MH, Liu P, Roth GA, Ng M, Biryukov S, Marczak L, et al. Global burden of hypertension and systolic blood pressure of at least 110 to 115 mm Hg, 1990-2015. J Am Med Assoc 2017;317(2):165-182 [FREE Full text] [doi: 10.1001/jama.2016.19043] [Medline: 28097354]
2. Neuhauser H, Kuhnert R, Born S. 12-Month prevalence of hypertension in Germany. J Health Monit 2017;2(1):51-57 [FREE Full text] [doi: 10.17886/RKI-GBE-2017-016] [Medline: 37151301]
3. Zhou B, Perel P, Mensah GA, Ezzati M. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. Nat Rev Cardiol 2021;18(11):785-802 [FREE Full text] [doi: 10.1038/s41569-021-00559-8] [Medline: 34050340]
4. Olsen MH, Angell SY, Asma S, Boutouyrie P, Burger D, Chirinos JA, et al. A call to action and a lifecourse strategy to address the global burden of raised blood pressure on current and future generations: the Lancet Commission on Hypertension. Lancet 2016;388(10060):2665-2712. [doi: 10.1016/S0140-6736(16)31134-5] [Medline: 27671667]
5. Parati G, Goncalves A, Soergel D, Bruno RM, Caiani EG, Gerdts E, et al. New perspectives for hypertension management: progress in methodological and technological developments. Eur J Prev Cardiol 2023;30(1):48-60. [doi: 10.1093/eurjpc/zwac203] [Medline: 36073370]
6. Kuwabara A, Su S, Krauss J. Utilizing digital health technologies for patient education in lifestyle medicine. Am J Lifestyle Med 2020;14(2):137-142 [FREE Full text] [doi: 10.1177/1559827619892547] [Medline: 32231478]
7. Mikulski BS, Bellei EA, Biduski D, De Marchi ACB. Mobile health applications and medication adherence of patients with hypertension: a systematic review and meta-analysis. Am J Prev Med 2022;62(4):626-634. [doi: 10.1016/j.amepre.2021.11.003] [Medline: 34963562]
8. Nationale VersorgungsLeitlinie Hypertonie (2023). URL: https://register.awmf.org/assets/guidelines/nvl-009l_S3_Hypertonie_2023-06.pdf [accessed 2024-04-04]
9. Bundesministerium für Gesundheit. Bundesgesundheitsminister Jens Spahn im Bundestag zum Digitale-Versorgung-Gesetz (DVG). URL: https://www.bundesgesundheitsministerium.de/presse/reden/dvg-23-lesung.html#c16594 [accessed 2020-12-15]
10. Dahlhausen F, Zinner M, Bieske L, Ehlers JP, Boehme P, Fehring L. There's an app for that, but nobody's using it: insights on improving patient access and adherence to digital therapeutics in Germany. Digit Health 2022;8:20552076221104672 [FREE Full text] [doi: 10.1177/20552076221104672] [Medline: 35811758]
11. Telematikinfrastruktur: alle Infos zur TI und ihren Anwendungen (2024). URL: https://abrechnungsstelle.com/telematikinfrastruktur/ [accessed 2024-04-04]
12. Müller T, Padmanabhan P, Richter L, Silberzahn T. e-Health Monitor 2022. URL: https://www.mckinsey.com/industries/life-sciences/our-insights/german-e-health-offerings-expand-but-adoption-remains-uneven#/ [accessed 2023-05-05]
13. Madanian S, Nakarada-Kordic I, Reay S, Chetty T. Patients' perspectives on digital health tools. PEC Innov 2023;2:100171 [FREE Full text] [doi: 10.1016/j.pecinn.2023.100171] [Medline: 37384154]
14. Morrissey EC, Casey M, Glynn LG, Walsh JC, Molloy GJ. Smartphone apps for improving medication adherence in hypertension: patients' perspectives. Patient Prefer Adherence 2018;12:813-822 [FREE Full text] [doi: 10.2147/PPA.S145647] [Medline: 29785096]
15. McBride CM, Morrissey EC, Molloy GJ. Patients' experiences of using smartphone apps to support self-management and improve medication adherence in hypertension: qualitative study. JMIR Mhealth Uhealth 2020;8(10):e17470 [FREE Full text] [doi: 10.2196/17470] [Medline: 33112251]
16. Bruch D, Muehlensiepen F, May S, Wengemuth E, Johannsen O, Reber KC, et al. Digital preventive measures for arterial hypertension (DiPaH) - a mixed-methods study protocol for health services research. Front Cardiovasc Med 2022;9:1089968 [FREE Full text] [doi: 10.3389/fcvm.2022.1089968] [Medline: 36704456]
17. Etikan I, Musa SA, Alkassim RS. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1-4 [FREE Full text]
18. Robinson RS. Purposive sampling. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Netherlands: Springer Dordrecht; 2014.
19. Hennink MM, Kaiser BN, Marconi VC. Code saturation versus meaning saturation: how many interviews are enough? Qual Health Res 2017;27(4):591-608 [FREE Full text] [doi: 10.1177/1049732316665344] [Medline: 27670770]

20. Stenzel R, Hadaschik K, May S, Grahammer M, Labinsky H, Welcker M, et al. Digitally-supported patient-centered asynchronous outpatient follow-up in rheumatoid arthritis - an explorative qualitative study. BMC Health Serv Res 2022;22(1):1297 [FREE Full text] [doi: 10.1186/s12913-022-08619-6] [Medline: 36307779]
21. Muehlensiepen F, May S, Hadaschik K, Vuillerme N, Heinze M, Grahammer M, et al. Digitally supported shared decision-making and treat-to-target in rheumatology: a qualitative study embedded in a multicenter randomized controlled trial. Rheumatol Int 2023;43(4):695-703 [FREE Full text] [doi: 10.1007/s00296-022-05224-y] [Medline: 36229671]
22. Kuckartz U. Qualitative inhaltsanalyse. 4th ed. Weinheim: Beltz Juventa; 2018.
23. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
24. Wangler J, Jansky M. Gesundheits-Apps in der hausarztbasierten versorgung ? Empirische befunde zur perspektive von allgemeinmedizinern und patienten. In: Pfannstiel MA, Holl F, Swoboda WJ, editors. mHealth-Anwendungen für chronisch Kranke. Wiesbaden: Springer Gabler; 2020.
25. Morrissey EC, Glynn LG, Casey M, Walsh JC, Molloy GJ. New self-management technologies for the treatment of hypertension: general practitioners' perspectives. Fam Pract 2018;35(3):318-322. [doi: 10.1093/fampra/cmx100] [Medline: 29088438]
26. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
27. Kassavou A, Wang M, Mirzaei V, Shpendi S, Hasan R. The association between smartphone app-based self-monitoring of hypertension-related behaviors and reductions in high blood pressure: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(7):e34767 [FREE Full text] [doi: 10.2196/34767] [Medline: 35819830]
28. Qudah B, Luetsch K. The influence of mobile health applications on patient - healthcare provider relationships: a systematic, narrative review. Patient Educ Couns 2019;102(6):1080-1089. [doi: 10.1016/j.pec.2019.01.021] [Medline: 30745178]
29. Mühlensiepen F, Kurkowski S, Krusche M, Mucke J, Prill R, Heinze M, et al. Digital health transition in rheumatology: a qualitative study. Int J Environ Res Public Health 2021;18(5):2636 [FREE Full text] [doi: 10.3390/ijerph18052636] [Medline: 33807952]
30. Muehlensiepen F, Petit P, Knitza J, Welcker M, Vuillerme N. Factors associated with telemedicine use among German general practitioners and rheumatologists: secondary analysis of data from a nationwide survey. J Med Internet Res 2022;24(11):e40304 [FREE Full text] [doi: 10.2196/40304] [Medline: 36449333]
31. Schaeffer D, Griese L, Haarmann A. [Professional health literacy of general practitioners - results of the HLS-PROF]. Gesundheitswesen 2024 [FREE Full text] [doi: 10.1055/a-2350-6377] [Medline: 39191374]
32. Labinsky H, Gupta L, Raimondo MG, Schett G, Knitza J. Real-world usage of digital health applications (DiGA) in rheumatology: results from a German patient survey. Rheumatol Int 2023;43(4):713-719 [FREE Full text] [doi: 10.1007/s00296-022-05261-7] [Medline: 36543961]
33. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014;39(5):356-364 [FREE Full text] [Medline: 24883008]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
DiGA: Digitale Gesundheitsanwendungen
HCP: health care professional
mHealth: mobile health
TI: telematics infrastructure

Edited by A Kushniruk; submitted 07.09.23; peer-reviewed by K Lakshminarayan, RJ Katz, YR Park, G Alnooh; comments to author 07.03.24; revised version received 25.04.24; accepted 31.12.24; published 10.03.25.

Please cite as:
May S, Muehlensiepen F, Wengemuth E, Seifert F, Heinze M, Bruch D, Spethmann S
Benefits and Barriers to mHealth in Hypertension Care: Qualitative Study With German Health Care Professionals
JMIR Hum Factors 2025;12:e52544
URL: https://humanfactors.jmir.org/2025/1/e52544
doi:10.2196/52544
PMID:

Original Paper

Perceived Risks, Mitigation Strategies, and Modifiability of Telehealth in Rural and Remote Emergency Departments: Qualitative Exploration Study

Christina Tsou[1*], BN, MPH, JD; Justin Yeung[2*], BMedSci, MBBS; Melanie Goode[2*], BSc; Josephine Mcdonnell[2*], BSc, MBChB; Aled Williams[2*], MBChB, MPH+TM; Stephen Colin Andrew[2*], BSc; Jenny Tetlow[2*], BEc; Andrew Jamieson[2*], MBBCh, MOccMed; Delia Hendrie[3*], PhD; Christopher Reid[3*], BA, MSci, PhD; Sandra Thompson[4*], BSc, MBBS, MPH, PhD

[1]WA Country Health Service, Curtin University, Perth, Australia

[2]WA Country Health Service, Perth, Australia

[3]Curtin University, Perth, Australia

[4]Western Australian Centre for Rural Health, University of Western Australia, Perth, Australia

[*]all authors contributed equally

Corresponding Author:
Christina Tsou, BN, MPH, JD
WA Country Health Service
Curtin University
Kent Street
Perth, 6102
Australia
Phone: 61 862661800
Email: tsouc01@gmail.com

Abstract

Background: Telehealth is a recognized and rapidly evolving domain in the delivery of emergency medicine. Research suggests a positive impact of telehealth in patients presenting for emergency care; however, the regional challenges of acute telemedicine delivery have not been studied. The WA Country Health Service (WACHS) established the Emergency Telehealth Service (ETS) in 2012 to provide telehealth and other technology-enabled services to regional Western Australian hospitals and clinics. The WACHS ETS supports 87 rural and remote WACHS-operated hospitals as well as 10 non-WACHS health clinics via high-definition audio-visual equipment installed in the resuscitation bay of the emergency department (ED) at each site. This 12-year practical application of emergency telemedicine offers a unique opportunity to explore the experiences and perceptions of clinicians delivering virtual care to rural and remote communities.

Objective: This study explores the perceptions of ETS clinicians regarding acceptability, appropriateness, and clinical decision-making when delivering emergency telemedicine in rural and remote settings.

Methods: This qualitative study used semistructured interviews to explore the perspectives of ETS clinicians regarding the factors influencing their clinical decision-making. It explored how ETS clinicians determine and modify clinical risks associated with using audio-visual equipment to deliver care. Emerging themes were compared with the concepts arising from the interim guidance of the Medical Board of Australia, and both the Australian and New Zealand, and American Colleges of Emergency Medicine.

Results: Overall, 16 doctors, 4 clinical nurse coordinators, and a nurse educator from WACHS ETS provided their experiences and perspectives. Accurate clinical decisions, especially regarding patient disposition, were crucial to virtual care. Timeliness and accuracy were enhanced through a mutual learning model grounded in the local context. Mitigation strategies such as improvisation and flexible technology use compensated for technological barriers. Nonmodifiable risk factors included patients' presenting complaints, clinical urgency of presentation, ED capability, clinician scope of practice, and, if a transfer was required, the distance between the ED of original presentation and the hospital of definitive care.

Conclusions: Telehealth can enhance clinical decision-making in rural and remote EDs, and ETS clinicians can prioritize patient safety through a lens incorporating both local hospital capabilities and community contexts. Even for the most experienced

clinicians, telehealth was not comparable to face-to-face communication in all circumstances. The impact of the ETS on the scope of the regional emergency medicine practice and on the building of clinical skills warrants further study in relation to its overall effectiveness and cost-effectiveness in rural and remote EDs. These findings identify areas for further qualitative research while providing a rich contextual background for rigorous quantitative analysis of the effectiveness of the ETS.

(JMIR Hum Factors 2025;12:e58851) doi:10.2196/58851

KEYWORDS

emergency telemedicine; implementation effectiveness; clinical effectiveness; risk aversion; risk mitigation; rural and remote; emergency departments

Introduction

Background

Telehealth has rapidly emerged as a means of providing health advice and care to people at sites where health providers at the hospitals of initial presentation (presenting hospitals) may lack capability or require assistance with the diagnosis and treatment of presenting clinical conditions. In recognition of the rapidly evolving domains of telehealth in the delivery of emergency medicine (EM) in the United States [1] and Australia and Aotearoa (New Zealand) [2], the Australasian College of Emergency Medicine (ACEM) [2], American College of Emergency Physicians [1], and Medical Board of Australia [3] deliberated on interim principles, considerations, and minimum standards to guide emergency telemedicine ("the interim guiding principles from the Colleges of Emergency Medicine"). Telehealth in the emergency department (ED) aspires to improve the quality of health care for patients in remote and other settings by increasing access to specialists and optimizing health service utilization and disposition coordination, including timely access to definitive care [2].

Guiding Principles on Emergency Telemedicine Implementation

These interim principles provide minimum standards for telehealth practice in the ED. As the Colleges of Emergency Medicine have identified, the use of telehealth in EM is in the formative stage and currently lacks a strong evidence base for clinical practice in the context of emergency care.

The ACEM and Australian New Zealand Fellow of ACEM (FACEM) Telemedicine Community of Practice (ANZFTCOP), superseded by a formal ACEM Emergency Telehealth Network, raised the importance of local context and the unplanned and exigent nature of emergency care [3]. The ANZFTCOP distinguished the nature of emergency medical practice from the broader telehealth discussion as it involves the provision of acute care without an ongoing clinical relationship either in-person or virtually [3]. The Community of Practice also pointed to circumstances where it may not be possible to access emergency physician expertise without recourse to telemedicine [3]. This is common for many rural and remote Australian EDs. The ANZFTCOP suggested that the decision regarding the appropriateness of conducting a telemedicine session instead of an in-person consultation should be based on individual clinical circumstances and expert clinician judgment [3]. Table 1 provides an amalgamated summary of the key guiding principles from these position statements and guidance documents.

Table 1. Amalgamated summary of the interim guidance on emergency telemedicine implementation.

Key themes	What is emergency telehealth and what should it do?	What is not emergency telehealth and what should it not do?
Locally grounded	• Should be an important complement for locally and regionally provided comprehensive health care. • Should be part of regional models of emergency care delivery validated by all relevant stakeholders across the public health system. • Should have regard for the local context and be grounded in the relationship with local clinicians including but not limited to knowledge of the local population, service availability, and patient pathways. • Should have appropriate escalation procedures in place to activate emergency services.	• Should not replace local care. • Should not be considered as a substitute for face-to-face consultations. • Should not involve low-value care that would not have otherwise been provided. • Should not be a stand-alone solution to address health workforce capacity and maldistribution.
Of similar quality and standard to ordinary care	• Assessments and consultations: (1) should be thorough within the limitations of telehealth and virtual consultations, (2) should replicate the components of an emergency medicine consultation, and (3) should be cognizant of the peculiarities or distinctive requirements of specific conditions. • Should implement quality assurance programs to monitor clinical performance, patient outcomes, and integrated ongoing care.	—[a]
Timely care	• Should ensure that definitive care is not delayed.	• Should not create additional barriers to accessing emergency medical care.
Appropriateness of care	• Should ensure that the patient's presenting problem is suitable to be assessed and managed remotely.	• Not appropriate for all clinical consultations.

[a]Not applicable.

Study Setting

The WA Country Health Service (WACHS) is one of the largest country health services in the world in terms of geographical coverage (2.55 million square kilometers, 96% of the land mass of Western Australia), with a population of 531,934 (18% of the Western Australian population; population density of 0.21 per square kilometer). It is organized into 7 distinct and diverse regions with 118 health facilities, accounting for 41% of the total number of emergency presentations across Western Australia (over 40,000 presentations per year) [4]. Most hospitals outside the large regional centers are staffed mainly by nursing staff supported by a resident general practitioner (GP) credentialed to provide medical services at the hospital. There are variable numbers of consultant specialists providing services at the larger regional resource centers. The majority of WACHS hospitals do not have on-site access to emergency medicine physicians (FACEMs).

The WACHS established the Emergency Telehealth Service (ETS) in 2012 to provide telehealth and other technology-enabled services to regional Western Australian hospitals and clinics. The ETS is the earliest and most established service stream of the WACHS Command Centre, which supports 87 WACHS hospitals and 10 additional non-WACHS facilities through high-definition videoconferencing services [5]. The Command Centre is a 24/7 virtual clinical and operational hub including services for general medical inpatients, mental health, midwifery, after-hours palliative care, and transfer coordination. It supports clinicians in regional hospitals and nursing posts by providing ready access to specialist clinicians who use technology, videoconferencing, and real-time data to assist in delivering quality patient care [6].

In addition to over 38,000 clinical consultations per year, the ETS supports rural and remote clinicians to maintain skills and knowledge through an education program and opportunities for professional development [5]. The consulting clinicians are located across multiple geographical areas (Perth, across regional Western Australia, other Australian jurisdictions, and internationally). Most medical practitioners are FACEMs complemented by EM-credentialed GPs and emergency nurse practitioners. The central hub of the Command Centre is based in Perth, and referrals are coordinated by a team of clinical nurses and clinical nurse consultants.

Referrals to the ETS vary across hospital types but are mainly from small hospitals (81.5%), nursing posts (9.7%), and integrated district health services (8.2%), and the ETS is least used by regional resource centers (0.7%) [4,7]. The proportions of ED presentations involving the ETS also vary across hospital types, with nursing posts and small hospitals involving the ETS in 21.3% and 23% of all ED presentations, respectively. Between July 2018 and April 2023, the uptake for nursing posts ranged from 2.8% for nonurgent presentations to 62.1% for emergency presentations that must be attended within 10 minutes. For small hospitals, the uptake ranged from 5.3% for nonurgent presentations to 47.9% for presentations requiring resuscitation. The more remotely located EDs appear to have greater reliance on the ETS especially for the management of more urgent cases [4,7].

Evidence on Telehealth Implementation in Rural and Remote EDs

Studies have identified factors that negatively influence the experiences of regional clinicians in rural and remote EM practices. These include [8,9] the lack or absence of medical backup or resources, the geographical and social isolation of

rural communities, poor health system coordination, the loneliness experienced by regional clinicians, and the challenges of health care provision in geographically isolated locations [10-12].

The challenges these conditions pose are community specific [8], and studies found that rural and remote EM physicians drew on the strength of professional relationships to improvise, solve problems, and create systems of support to meet local needs in order to survive and thrive in the middle of rural and remote challenges [8,9].

Published studies on the use of telehealth in EM suggested positive impacts on patient care, including benefits such as cost reduction, improved quality of care, reduced mortality rate, reduced patient treatment time, and reduced time between first contact and treatment [13]. Moreover, in the rural and remote context, studies have reported reduced patient transfers from rural facilities to major centers [10-12,14-16] and improved capability of rural centers [10,17,18]. However, studies have not discussed the complexities of rural and remote EM practices in depth or taken them into account in their analyses.

The use of telehealth in rural and remote EDs to help alleviate these regional EM practice challenges has not been studied. Whether or how emergency telemedicine in rural and remote settings improves patient health outcomes and the factors influencing the clinical decisions of emergency telehealth clinicians remain uncertain. There is a need for quality evidence to better understand these factors to help identify the gap in the effectiveness of emergency telehealth [12,14,15,17,19-27].

Studies have, however, identified challenges in the implementation of virtual care, and emergency telehealth is one of the practice areas considered [14,22,23]. These include technical difficulties [11,12,23,25-27], factors impeding the process of care [10,14,16,19-21,24-26], and factors impacting patient privacy, confidentiality, and data security [24-26].

It has been proposed that to understand the factors influencing job satisfaction, the complex environments around rural and remote EM practices must be considered [8]. However, the impact of environmental factors on clinical decision-making in rural ED practice has not been considered.

The clinicians' perceived clinical risks have been identified as a challenge in rural and remote EM practices [8,9]; however, how individuals perceive the risks in rural and remote emergency telehealth practices and the impacts on their clinical practices have not been explored previously. In rural and remote EDs, it is possible to assume that when the perceived clinical risk is high, the approach to diagnosis and treatment may change, leading to increased rates of transfer. The distinctions between these perceived clinical risks when delivering face-to-face versus virtual EM services warrant investigation.

To the best of our knowledge, specific links between the challenges of clinical reasoning and decision-making in virtual emergency care and the rural and remote context have not been identified. However, these factors are critical for understanding the effectiveness of implementing telehealth in rural and remote EDs.

Study Aims

From the perspective of WACHS clinicians delivering the ETS, this study explored the perceptions of emergency telehealth clinicians regarding the acceptability, appropriateness, and determinants in clinical decision-making when delivering emergency telehealth in rural and remote settings. This study aims to understand the following: (1) the factors influencing clinical decisions during an emergency telehealth consultation; (2) the impact of involving emergency telehealth clinicians in rural and remote ED consultations on the health workforce and the safety of clinical service delivery; and (3) how the clinical risks associated with emergency telehealth consultation can be modified, including whether improving the status of factors that are amenable to change can positively influence the safe and effective delivery of the ETS.

Methods

Overview

This qualitative study used semistructured interviews to explore the perspectives of ETS clinicians regarding the factors influencing their clinical decision-making. It explored how ETS clinicians determine and modify associated clinical risks. Emerging themes were compared with the concepts arising from the interim guidance of the Medical Board of Australia and both the Australian and New Zealand, and American Colleges of Emergency Medicine. The WACHS ETS was used to examine contextual factors influencing the quality and perceived effectiveness of emergency telehealth in rural and remote areas.

Sampling Techniques and Study Participants

A combination of convenience, snowball, and purposive sampling was used to obtain a diverse perspective on the ETS model of care. Convenience sampling was required given the nature of ED work and participant availability to recruit FACEMs, GPs, ETS nurse coordinators, and clinical nurses. A purposive sampling of WACHS ETS physicians who were interstates or overseas occurred toward the end of the participant recruitment.

On December 17, 2020, an invitation email was sent to all ETS doctors and nurses by one of the FACEM members of the research team. This was followed by individual follow-up emails on January 22, 2021, to ETS doctors, as well as emails and in-person follow-ups by one of the ETS clinical nurse consultants in the research team.

Interview Guide Development

The interview protocol (guide) was developed in collaboration with the clinical researchers at the Command Centre. This interview guide was then distributed to the full research team for review.

After piloting the interview protocol with a select leadership group, the protocol was updated to include probes on whether an electronic stethoscope would help provide confidence in clinical decision-making, as well as adding questions related to the clinical governance, service legitimacy, and legality of the ETS practice. Minor changes were also made to the wording and ordering of the content areas and probes within each content

area. The final version of the interview protocol is presented in Multimedia Appendix 1.

Simultaneous Data Collection and Analysis

Semistructured in-depth interviews explored the experiences and opinions of the clinicians involved in ETS implementation and the factors considered in clinical decision-making.

Participants were given the option to participate in an interview via MS Teams (Microsoft Corp) or phone, from a place and at a time of their choosing. Interviews were audio recorded, automatically transcribed, and repeatedly played back in the open-coding phase of the analysis to validate the accuracy of the transcription. Due to the nature of emergency work, the recruitment and conduct of the interview occurred opportunistically to catch moments when the clinicians were available. This meant that interviews sometimes occurred outside rostered work times, with some conversations interrupted by work-related demands.

Data analysis proceeded simultaneously with data collection and issues arising from information [28]. The early phase of the analysis included open and axial coding [29], identifying the emerging themes with open coding from clinician interviews. The emerging themes were then compared with the guiding principles identified from the interim guidance documents of the Colleges of Emergency Medicine and Medical Board of Australia. The key guiding principles were introduced as axial codes to place the open codes within the context of EM practice. In the final phase of the analysis, risk factors considered by clinicians in their clinical decision-making during an emergency telehealth consultation were extracted and classified into technology related, presenting hospital/location related, or EM related. The modifiability of each of the factors was assigned by the lead author and confirmed by co-authors based on author experience, the rural and remote context of Western Australia, the public health service landscape, and the resources available to the WACHS. The factual findings reported in this paper were discussed with and confirmed by the WACHS Command Centre.

Ethics Approval

This research obtained ethics approval from the WACHS Human Research Ethics Committee (approval number: RGS0000003076) and reciprocal ethics approval from the Curtin University Human Research Ethics Committee (approval number: HRE2019-0740-09). Site access was governed by the WACHS research governance policy and procedures. A participant information and consent form was attached to individual email invitations to the participants to review, sign, and return before the scheduled interview. For participants who attended the interview but were unable to sign the consent form, verbal consent was obtained before the commencement of the interview. Interview recordings were transcribed verbatim by the first author. The transcripts and recordings were allocated a participant code and stored in a password-protected electronic folder that is separated from the folder where the participant register is stored.

Results

Interviews and Participants

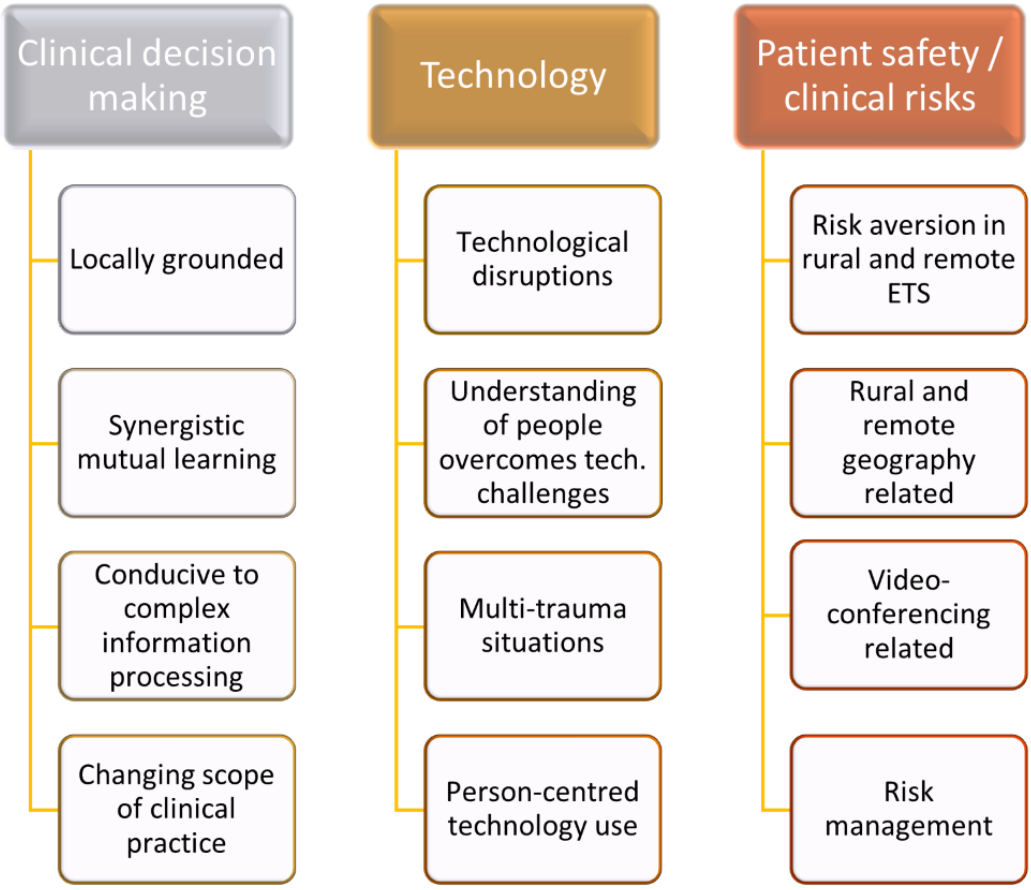
Most interviews were conducted via MS Teams (17/21, 81%), and the remaining were conducted by phone (4/21, 19%). Participants were in their private homes, a dedicated home office space, or Western Australian health facilities.

Sixteen WACHS ETS doctors and 5 ETS nurses participated in the interviews. There were 9 FACEMs, 5 GPs, 2 ED registrars, 4 clinical nurse coordinators, and 1 nurse educator. All participating ETS nurses were based at the Command Centre in Perth, while 5 of the ETS medical practitioner participants were based in Perth, 6 were based in regional Western Australia, 4 were overseas, and 1 was interstate. Two senior rural GPs provided their perspectives from both the receiving and provisioning ends of the ETS (as GPs working in regional locations). Around a quarter of ETS doctors and 16% of ETS nurses in the Command Centre participated in the interviews. All ETS nurses, except 1, were recruited by the same clinical researcher through face-to-face invitations, while 83% (15/18) of the ETS doctors who participated in the interviews responded to the initial invitation to participate. The purposive recruitment yielded 2 additional ETS participants to add weight to the perspectives of overseas or interstate doctors. Multimedia Appendix 2 lists the details of the participants, including home base, professional designation, gender, ETS role, and perspectives they contributed to this study.

Domains That Emerged From the Simultaneous Data Collection and Analysis

The major thematic domains identified in this analysis were clinical decision-making, technology, and patient safety or clinical risk. The themes and subthemes have been outlined below. Figure 1 provides a representation of the domains and themes.

Figure 1. Thematic domains and subthemes that emerged in this study. ETS: Emergency Telehealth Service; Tech.: technology.



Clinical Decision-Making: Benefits of Involving the ETS to Support Clinical Decisions

Participants identified 3 key benefits of ETS involvement in rural and remote clinical services. Hospital clinicians reported that having the ETS improved the willingness and confidence of regional clinicians to practice in remote locations. The ETS enabled synergistic mutual learning through the collaborative patient management of site-based and ETS clinicians. Further, ETS FACEMs experienced the benefit of ETS practice over site-based practice as it allowed them to be more cognitively focused when processing complex information. Participants touched on the changing scope of clinical practice as part of their discourse on clinical decision support.

Locally Grounded Clinical Decisions

Participants described timely and appropriate disposition (admission, discharge, or transfer) of patients as the most significant clinical decision for rural and remote EDs. Disposition of ED patients, especially from less-resourced remote EDs, has significant patient safety and clinical risk implications. Clinical decisions pertaining to patients presenting to small and remote EDs differed from those of better-resourced urban or regional center EDs as site-related factors needed to be considered.

All participants commented that multiple interacting factors were at play when deciding to transfer a patient for ongoing management. Table 2 lists the factors they considered in their clinical reasoning, and only 1 was related to technology. The predominant consideration was the resources available locally within the presenting hospital and in the local community. The next consideration was the distance to a destination site of definitive treatment if a patient's condition deteriorated. The final consideration was the timing of transfer and its appropriateness according to an individual patient's health condition.

Table 2. Thought processes in clinical decisions including patient disposition considerations via rural and remote emergency telehealth.

Thought processes of decision-making clinicians	Technology related	Hospital/location related	Patient related	Emergency medicine related
Normal diagnostic decision-making: what is wrong with the patient?	No	No	No	Yes
Can the patient be adequately assessed for all relevant possibilities at the site? What [differential diagnosis] am I potentially missing and is that putting the patient at risk?	Yes	Yes	No	No
Is this patient safe to go home medically and socially?	No	Yes	Yes	No
Is there a local GP[a] in town to care for the patient the next day?	No	Yes	No	No
Am I happy for the patient to wait till the next day for follow-up in a community?	No	Yes	Yes	Yes
Considering the patient's location, the weather conditions, and the local resources available, is the patient safe to stay at this site and/or be admitted locally for observation?	No	Yes	No	No
Does the patient need to be retrieved earlier [compared to the standard of a tertiary center or regional center ED[b]]?	No	Yes	No	No
Where and how do I transfer them? Is this patient safe to go somewhere with private transport or road ambulance, or do they require aeromedical transfer?	No	Yes	Yes	No
What does the patient want? (some patients do not want to leave their community)	No	No	Yes	No

[a]GP: general practitioner.

[b]ED: emergency department.

Synergistic Mutual Learning With Joint Clinical Decision-Making

ETS physicians were reliant on the local knowledge of hospital clinicians at the site of patient presentation to facilitate and consider community, social, and resource factors in clinical decisions for regional patients. Rural-based hospital clinicians reported benefits from the assistance of ETS clinicians in making clinical decisions, especially their support in connecting to appropriate and available resources.

Regional-based participants reported that the involvement of ETS colleagues supported them in answering the questions listed in Table 2, with increased speed and confidence. An ETS physician who also worked at a site based on a WACHS ED stated:

> *It makes certain that we make the best decisions for our patients. ETS does make the decision-making process much easier in the region. I've got somebody else online who can call me if I'm looking after somebody who's critically ill. They can help organize transfers and ICU admission, they can help me with drug doses, so I can…crack on and do what I need to do. And I know that I'm working in a safe environment.* [ETS Doctor #3]

Participants also stated that they acquired a different skill set working in a supportive role to clinicians in remote facilities, including ways of communicating instructions to presenting hospital clinicians (mainly nursing staff), gathering medical history, prescribing medications, and efficiently analyzing and synthesizing presenting complaints.

For ETS clinicians who were FACEMs mainly based in metropolitan hospitals, participation in ETS consultations provided a new and unfiltered experience in the rural and remote clinical practice. Many reported that this changed the way they practiced in their substantive role on the floor of tertiary EDs:

> *I think [the learning has been] in both directions, one, it makes me realize how resource rich we are in a tertiary hospital. Two, it makes me realize that the nurses and other staff out there [in regional WA] are often just as competent, just as knowledgeable, and just as skillful and often braver than we are in tertiary hospitals. It certainly helps me stay grounded when I'm accepting referrals. [Tertiary hospital] has a big catchment, we receive a lot of these people from peripheral hospitals and nurse run clinics. And one that certainly helped me know where all these places are, when I'm on the receiving end of the phone…I'm able to explain to other staff exactly how resource limited, a lot of these clinics are. And there is a reason they haven't had blood tests and haven't had imaging. And in fact, there's no doctor there. So they have just been seen by nurses must come to be seen by us, we don't have to be difficult about accepting them. I think that's been quite useful.* [ETS Doctor #43]

Conducive to Complex Clinical Information Processing

Participants reported that their ability to control and prioritize their consultations in the virtual environment assisted their efficiency in processing complex information and enabled more objective assessment than working in person on the ED floor, given their distance from the patient. Participants reported the

benefits of providing virtual care compared to working on an ED floor:

> *If you're on an ED floor you are interrupted constantly because you're physically there…one of the most frustrating things is you're constantly distracted, and you can't finish a task from A to B, I think that adds hugely to burnout. …But you are focused here [at ETS], …we can only see one camera…one patient at a time. …And the chaos is more manageable… you can choose to not be interrupted. So I feel that it's very satisfying, because you're so cognitively focused. And I think that can potentially lead to better quality care… because you can just make better decisions.* [ETS Doctor #6]

With the ability to control the environment they practice in, ETS FACEMs found the virtual environment conducive to them being more cognitively focused.

Changed Scope of Clinical Practice

The impacts on the scope of practice of regionally based clinicians, especially nurses, were dichotomously reported. Most participants reported an increasing scope of site-based clinical practice with ETS doctors guiding local nurses to perform procedures that they otherwise would not have the confidence to do on their own. A contrary perspective was the engagement of ETS clinicians in treatment and procedures or clinical decisions, which local clinicians would have performed independently through a local process in place prior to ETS implementation. For example, regarding giving a tetanus booster, 1 participant commented:

> *…we'll get consults of patients who need a tetanus booster, they've had a small wound a few days ago, …I don't know what they would have done before, I think they would have done it and had a process in place to do it.* [ETS Doctor #2]

The changing scope of practice was noted by the participants. This reflected policy changes to increase medical practitioner involvement in ED consultations and was an emerging theme that warrants further in-depth exploration.

Technology

The second thematic domain was technology, which included technological (digital) disruption and mitigation strategies in emergency telehealth consultations.

Participants were of the view that the WACHS ETS had already made most of the technological and process improvements and considered that any further changes on the technological front would have incremental benefits in effectiveness for improving patient health outcomes. One FACEM participant stated:

> *A lot of the system problems are the fact that we have a hundred places that are tiny and it's hard to get [for example] a pathologist, or experienced nurses for long stays. A lot of it is system problems that are difficult to modify.* [ETS Doctor #43]

> *We probably made the easy gains and the incremental changes you can make to improve things from here are big and costly, and difficult. For example, attracting and keeping people in remote communities and providing enough staffing to let GPs take holidays and do professional development and be replaced by someone, rather than leaving the community to rely on telehealth because there is no one else to cover.* [ETS Doctor #43]

Participants described the ways in which they modify how they use the telehealth technology or adjust the process of care to improve the quality of ETS consultations.

Technological (Digital) Disruptions

A commonly expressed view among ETS physician participants was that during videoconferencing, one is expected to lose nonvisual cues and some verbal cues due to technological (digital) disruptions. Clinicians reflecting on the videoconference consultation experience suggested that the amount and quality of information a clinician could gather through video came down to their level of understanding of people as individuals. They referred to the nuances of patients' expressions observable during videoconferencing and being able to relate to the regional context of the patients. Table 3 provides a summary of the circumstances around the technological barriers during an ETS consultation and the mitigation strategies participants used to overcome these technological disruptions.

Table 3. Technology-related issues of Emergency Telehealth Service technology and mitigation strategies.

ETS[a] technology-related issues	Technology barrier			Mitigation strategy		
	Barrier	Circumstances	Participants	Strategy	Incompatibility	Participants
Positioning of the monitor where the image of the ETS physician appears (ie, at the foot of the ED[b] stretcher)	• Sending the wrong message that the doctor is looking down on the patient. • Abrupt appearance does not allow ETS to ease into the scene as in the case of an in-person ED floor.	When people on the floor are not adequately prepared and people are unaware why the ETS clinician is appearing on the screen, there can be disruptions in the actions already taking place in the ETS bay.	ETSD[c] #1, 4, 6, 7, 10, and 13; ETSN[d] #3	• Prepare patients before the VC[e] link is up; explain the process before and after VC consultation, including the steps leading up to transfer when interhospital transfer is indicated • ETS physician is identified as not from Perth and thus does not appear as "someone from Perth talking down on us"	• In a critical encounter, there is no time and space to prepare patients and people in the room. • Only available to regional-based ETS clinicians.	ETSN #3; ETSD #7
Talking into an open space; broadcasting of sound from the VC setup into the ETS bay	• Impersonal, with privacy and confidentiality potentially compromised; concurrent private conversations can occur from adjacent ED bays, which can create distracting noise during the consultation process.	Especially when multiple parties are involved in the conversation	ETSD #1, 4, 7, 10, and 13; ETSN #3	• Mute VC sound and use a telephone directly with site-based clinicians or patients for audio to ensure private conversations. • Transition to telephone only after an initial meet and greet with VC. • Clinicians on the floor are encouraged to wear headphones to enable switching voice channels between different target audiences.	• If children are involved, communicate with parents or carers	ETSN #3 and 4
The ETS clinician's visual field of the ETS bay is restricted by the camera's visual field	• ETS clinicians are not aware of others in the room.	Consulting before realizing a family member or another patient is in the room; knowing who is in the room and their relationship with the patient can change the dialogue and alter the instructions given regarding the next steps.	ETSD #1 and 6	• Use a telephone before screen appearance, find out who is present, and clarify the purpose of involvement. • Self-introduction, role of ETS consultation, and a name shown at the base of the screen.	—[f]	ETSD #2, 4, 8, 9, 12, 13, 15, and 43
Zoom in or out function adjusted locally without the ETS clinician's awareness	• Unable to zoom in to see what needs to be seen.	Significant (disastrous) impact in critical situations.	ETSD #5	• Having a reliable camera that the ETS doctor can control.	—	ETSD #4 and 14

[a]ETS: Emergency Telehealth Service.

[b]ED: emergency department.

[c]ETSD: ETS doctor.

[d]ETSN: ETS nurse.

[e]VC: videoconferencing.

[f]Not applicable.

The videoconferencing equipment setup and the restriction of the camera's visual field created technological barriers affecting the quality of ETS consultation. Physical distance was the perceived distance between either side of the high-definition audio-visual equipment installed in the resuscitation bay of the ED at each site.

Participants experienced constraints of the camera setup made consultations impersonal and visual cues easily missed or distorted, and they shared how they adapted in their ETS practice to ensure effective visualization. One doctor said the technical issues and logistics of commencing a consultation with sudden appearance on a screen and sound broadcasting as key differences to working on the floor:

> *…, you got to spend time to dial in, and, then suddenly just appear on the screen in front of the patient… sometimes they're a bit shocked because you're big and loud on the screen… especially if they're drunk, confused that can exacerbate that [shock]. And sometimes I don't know what the volume is, sometimes I'm absolutely booming. And then everybody can hear me in the ED… they got to find the mute button, unmute me and decrease the volume…* [ETS Doctor #6]

A few participants suggested that adding a reliable remote-control pan or zoom would considerably improve the quality of the ETS physical assessment:

> *…great history and reading the cues if someone is lying very still, are they squirming around, are they making reasonable eye contact, are they coherent when they talk, all of that sort of stuff... you know looking for facial droop, obvious distal neurological deficit can be done really well [with reliable and high resolution camera].* [ETS Doctor #14]

Understanding of People Can Overcome Technological Challenges

Through reflective practice, participants redefined and modified their practice over the ETS to mitigate the shortcomings of the videoconference equipment setup. There were varied responses to technological (digital) disruptions in interpersonal communications. ETS doctors gave examples of how they used visual and imperfect auditory cues to assist with formulating their working diagnosis when direct eye contact was not possible via videoconferencing. Visual cues, such as closing hands and fidgeting, were used to the clinician's advantage to help with diagnosis. Humor and a nonjudgmental approach were helpful in getting patients to reveal their alcohol issues and to work collaboratively with patients over videoconferencing to reach the best outcome for them. These human approaches made a notable difference in the virtual physical assessment process.

Multitrauma Situations With One Videoconference Bay

In a multitrauma situation when only 1 doctor was available on site, teamwork was required. For example, when the site-based doctor was busy with hands-on procedures in collaboration with an ETS FACEM, a local nurse stepped up to perform the team leader tasks under the guidance of ETS clinical nurses. When there were multiple casualties, all casualties would need to rotate to be in front of the cameras (typically 2 perpendicular pan/tilt/zoom cameras), and by using headphone channel switching, ETS clinicians could converse with each of the site clinicians individually. When there was a requirement to "shuffle" beds in and out of the camera visual field, adequate explanation ahead of the action was noted as an indispensable component of patient-clinician communication.

Person-Centered Technology Use

Facing variations in the quality of the internet connection, fixed positioning of cameras, and the broadcasting of sound into an open space of the ETS bay at regional EDs, ETS clinicians developed approaches to work with and via the presenting hospital clinicians to improve the quality of videoconference consultation while also protecting the privacy and confidentiality of patients. Overcoming technological limitations required users to troubleshoot and adapt to circumstances, requiring flexibility to ensure the effectiveness of virtual care. This finding pertains more to the exigent nature of EM and less to the rural and remote context of ETS delivery.

Patient Safety and Clinical Risk Considerations

The final thematic domain was related to the perceived risks of virtual care identified by the study participants (Table 4). The central aim was to ensure patient safety when using telehealth. If the ETS team perceived that patient safety was compromised by remaining at the local facility (particularly if there was no on-site medical practitioner), they would strongly consider transferring patients to an appropriate hospital for ongoing management. ETS clinicians who had higher levels of perceived clinical risk had a lower threshold to transferring their patients. Participants were confident that with experience (both at the site and virtually), some transfers could be avoided. The modifiability of risks as assessed in this study identified areas for potential improvement and the factors to consider in the effectiveness dialogue.

Table 4. Factors that make emergency telemedicine clinicians risk averse in rural and remote areas in Western Australia.

Clinical risks and sources of risks in the Emergency Telehealth Service practice	Technology related	Hospital/location related	Emergency medicine related	Is it modifiable?
Insufficient or lack of local resources				
Nature of ED[a] cases: sudden deterioration not always predictable	No	No	Yes	Not modifiable
Distance/time required to transfer to a location with more resources against the risk of deterioration	No	Yes	No	Not modifiable
No short-stay unit: cannot keep the patient after hours	No	Yes	No	Potentially
Local workforce capacity and capability				
Clinicians on the ED floor are unable to perform the required physical assessment	No	Yes	No	Limited
Procedural (IV[b] access and treatment procedures)	No	Yes	No	Potentially
No doctor or insufficient nurses onsite to admit the patient	No	Yes	No	Potentially
No GP[c] in town to follow-up on the next day	No	Yes	No	Limited
Unable to secure a clear diagnosis				
Unable to access required testing or imaging	No	Yes	No	Not modifiable
Unable to fully examine: chances of missing a clinical cue are greater	Yes	No	No	Limited
Not 100% confident with the clinical picture despite full complement of clinical information	No	No	Yes	Not modifiable
Variable internet quality: disrupts physical assessment	Yes	No	No	Limited
Camera is not always reliable	Yes	No	No	Modifiable
Individual factors				
Physician personality and approaches to risk	No	No	Yes	Not modifiable
Physicians are less comfortable with the reliability of their own physical assessment over VC[d] compared to face-to-face evaluation	Yes	No	No	Potentially
Patient preference (to stay closer to home)	No	No	Yes	Not modifiable
Patients and family members are risk averse to telehealth consultation	Yes	No	No	Potentially

[a]ED: emergency department.

[b]IV: intravenous.

[c]GP: general practitioner.

[d]VC: videoconferencing.

Concept of Risk Aversion in the WACHS ETS

Risk aversion impacts transfer rates from the site of presentation. The level of risk ETS clinicians were willing to take in keeping patients locally depended on 4 broad categories risk considerations. The first and second considerations of risks were related to the presenting hospital capacity and capability (ie, the health service resources and workforce). Risk aversion was associated with increasing remoteness of the presenting hospital. The third category of risk was unanimously reported by all the ETS physician participants. They reported that they were not able to examine the patient adequately or be completely confident with the clinical information relayed by site-based clinicians. This concern reflected how well ETS clinicians felt they could apply strategies to mitigate technological (digital) disruptions. The final category reflected individual clinician factors, such as their personality and their own tendency to be risk averse, and was present in EM generally, irrespective of whether it was place based or virtual. Factors related to the nature of EM and patients presenting to EDs would not have changed this risk even if managing clinicians were able to be "teleported" to the scene in person, and these aspects have not been discussed further in this paper.

Risk Aversion Associated With the Hospital of Presentation and the Rural and Remote Geography

Of 17 risks, 8 emerged from the ETS risk-aversion discussion related to the hospital of presentation and the rural and remote geography. These were less readily modifiable than the videoconference-related risks. The least modifiable risk was related to the inability to access the required tests or imaging, and this was reported by multiple clinicians. If point-of-care testing and the pathological approach available locally were not adequate to reach a clear diagnosis, the patient would be

transferred for further investigation. The extent to which this was considered an issue varied, with 1 participant (ETS doctor #3) having a view that point-of-care tests, routine pathology, and x-ray imaging available locally were generally satisfactory and sufficient for many of the presentations, and this view differed from that of many other participants.

Another factor associated with the presenting health facility was related to the regional workforce capacity and capability to perform the required physical assessment or procedure, or to keep patients locally for observation onsite or in the community. This mainly applied to sites where there were no medical practitioners. These factors had limited modifiability at the time of the consultation but were potentially adaptable. When the ETS clinician was unable to work with the clinician onsite to overcome the limitations of physical assessment over videoconferencing, the workforce capacity at the presenting hospital impacted the ability to secure a clear diagnosis and manage the patient adequately. Workforce capability was also related to the inability to perform required procedures ranging from establishing vascular access, inserting an intraosseous needle, and simple suturing to performing more complicated procedures, which could have avoided transfer. Generally, this is related to nursing-only facilities staffed by clinicians with limited scope of practice. Even when a diagnosis was made and the required procedures are performed, a patient might still be transferred if there was no doctor or insufficient nursing capacity to keep the patient on site for further observation, including if there was no GP in town the next day to follow-up in the community.

The final factor adding to the risk aversion of ETS clinicians was related to resources available at the hospital of presentation. The key issue raised here was the absence of a short stay–type inpatient unit to keep patients in the hospital for overnight observation. The distance and time required to transfer patients to a location with higher levels of capacity and capability were considered against the risk of deterioration of the patient's condition and the assessment of clinical risk. These factors were not modifiable and subject to the treating clinician's judgment of the patient's health condition.

Videoconference-Related Risk Aversion

Of the 4 categories of ETS clinical risk, 2 included factors that were related to the inability to secure a clear diagnosis and were related to videoconferencing. Many participants discussed the need to transfer patients because they were not confident with the clinical assessment via video. These included their perception of missing significant clinical signs, due to being unable to directly palpate or examine the patient, which impeded reaching a clear diagnosis. For example, 1 FACEM participant commented as follows:

> *…spending a lot more time thinking about exactly what I'm going to do and why… you don't have any of the nice reassurance normal investigations you would normally have. You are putting yourself more at risk by doing telemedicine because the chances of missing something are greater because you don't have access to all the little things that would make you much more comfortable.* [ETS Doctor #14]

ETS physician participants also reported issues with internet connectivity and speed, and many made comments on issues associated with the reliability of the camera affecting their ability to formulate a clear diagnosis. As discussed above, issues related to the camera were associated with the way it was physically positioned in the allocated treatment bay; operational limitations with pan, tilt, or zoom; and the camera technology itself. The internet speed and connectivity were more challenging as they were impacted by the internet provision where both the patients and ETS clinicians were physically based (nationally and internationally).

Finally, user acceptance of videoconference consultation differed among participants. Not seeing patients face-to-face made some clinicians more cautious and risk averse, with treating clinicians also reporting that some patients or family members were risk averse to telehealth consultations due to a lack of experience or understanding of this alternative modality of clinical service delivery. Several ETS clinicians reported that they spent more time second-guessing their observations and decisions, suspecting that they are less reliable over videoconferencing compared to a face-to-face meeting.

Risk Management for Patient Safety Instead of Risk Aversion

Complementing the higher level of potential and perceived risk associated with videoconference consultations, participants shared a range of perspectives on the perception of clinical risk:

> *…am I risk averse? I am probably less. It's not my personality. My personality is not risk averse. And I think most emergency physicians are not risk averse. That's not the nature of our business. There's risk everywhere we are risk managers.* [ETS Doctor #2]

Despite observing the potential risks associated with videoconference consultations, participants believed that those who were overly risk averse would not participate in telehealth, and it was a matter of finding a balance between the perceived benefits of ETS and the expected risks associated. A pragmatic view was that ETS physicians were no more cautious on camera compared to being on the ED floor; however, due to incomplete data points to inform an evidence-based clinical decision, interhospital transfer may be suggested for the required investigations. In terms of confidence in their decisions, some participants reported that they felt comfortable with their clinical decision-making as FACEM training helped balance relatively scant information in undifferentiated patients, and this was no different from the decision they would make in a face-to-face meeting:

> *Well, that's my job. It's not so much about being confident. Well, you've got to be confident in your decision because otherwise you can't afford to be in the [emergency medicine] business. I mean, you know, if you if you're worried about the patient, then you do something to alleviate that worry… Occasionally, there are times when it is difficult… for example, if a patient has a communication problem in either they are deaf, or they can't speak properly, and so my interaction is sufficiently degraded but otherwise, for*

> *[most] times, I don't think it does make a particular difference to be honest. The difference between me not being there and being there is far more evident in the skills that are brought face-to-face. And those are nearly all procedurals.* [ETS Doctor #5]

Discussion

Overview

The experiences and perspectives of WACHS ETS clinicians contributed to an in-depth understanding of the factors influencing clinical decision-making during an emergency telehealth consultation in rural and remote settings. The study progressed the understanding of the mutual learning opportunities presented during the joint decision-making process, the perceived clinical risks of emergency telehealth practice in the rural and remote context, and the technological barriers imposed on clinical decisions and the modifiability of these perceived risks.

Principal Results: Meeting the Principles of the ETS Practice

This research described the mutually beneficial effects of ETS consultation recognized by the participants in their clinical practice. These were in line with the principles and interim standards stipulated by the Colleges of Emergency Medicine that telehealth should complement and be part of the regional model of emergency care delivery validated by all stakeholders [1-3]. This finding also aligns with previous research, which highlights how rural and remote ED clinicians rely on the support and expertise of their colleagues to navigate and succeed in the unique challenges of rural and remote settings [8-10,17,18]. It further explored how this dynamic operated when colleagues provided support through a virtual presence. This research provides additional insights into how collegiality during emergency telehealth consultations supports various aspects of practice. These include enhancing confidence in rural and remote clinical practices, advancing the practice of virtual EM, and incorporating the rural and remote context into collaborative clinical decision-making involving stakeholders on both sides of the technology.

Clinical Decision-Making

The participants described clinical decision-making mediated by a complement of ordinary technologies. Considerable regard was given to the context of the hospital of initial ED presentation, the community where the patient is from, and the regional Western Australia health service delivery landscape. ETS clinicians were reliant on the local knowledge of presenting hospital clinicians, the patients, and their family members to incorporate these local contexts into collaborative decision-making. The capacity building was mutually beneficial for regionally based clinicians and ETS clinicians. Their complementary skills and synergies in what they were able to contribute to patient care became apparent during ETS consultations. In this paper, we have referred to this phenomenon as “synergistically mutual learning.”

Many participants described their experience of improved efficiency in processing complex clinical information through better focus. This reflected the time-sensitive nature of the ED, especially when the hospital of definitive treatment was hours away, and the organized chaos in multitrauma situations. This finding represents a novel contribution from the perspective of emergency physicians conducting virtual ED consultations. The quality of clinical reasoning was reported to be superior in ETS practice where there was greater reliance on synthesizing medical history, reviewing clinical records, and carrying out focused processing of multiple sources of information. In situations where there was a heavy reliance on physical examinations and complex laboratory or imaging investigations, clinicians were cognizant of the peculiarities and distinctive requirements of the specific conditions and replicated, as much as possible, the components of an EM consultation. This was in line with the guiding principles stipulated by the Colleges of Emergency Medicine to deliver a service of similar quality and standards as ordinary care [1,2].

The perspectives of ETS clinicians highlight a critical aspect of improvement in the overall efficiency of emergency care delivery not previously reported in studies of the efficiency of emergency telemedicine within receiving hospitals [10-12,14-16].

Clinical Use of Technology

At the forefront of technological change is how humans personalize technologies. The way technologies are used is a significant part of the innovation in the EM practice in rural and remote settings. This study extended past findings of technological difficulties impeding care processes [10,14,16,19-21,24-26], patient privacy, confidentiality, and data security [24-26] by explaining how emergency telehealth practitioners introduced human factors to the innovative use of emergency telehealth through their understanding of people and tailoring the use of the same technology to the needs of patients on the other side of the camera as described in Table 3. To the best of our knowledge, this is the first description of how human users innovate to close the distance across technological barriers in an emergency telehealth scenario.

Modifiability of Patient Safety and Clinical Risks Identified

Risk factors related to ETS technology were modified through upgraded camera resolution, fixed camera positioning, increased user experience, improvisation, and consumer preparation. The quality of physical examinations was generally considered unlikely to match the quality of face-to-face examinations, especially when examinations of hard-to-visualize areas of the body or procedural interventions were indicated. However, emergency telemedicine using videoconference facilities has been considered to have reached a plateau and might be sufficient to cover a great majority of presentations. Additional investment in other examination equipment will need to be carefully evaluated as the benefits of further equipment upgrades are likely to be incremental against the costs of purchasing the equipment.

To meet the standard of replicating components of an EM consultation stipulated in the interim guidance [1,2], the need for an appropriately skilled and competent local workforce

cannot be overlooked. Workforce capabilities were potentially modifiable through education, including pre- and postvocational training opportunities; however, modifiability was offset by the ongoing issue of the turnover of clinical staff in rural and remote hospitals. This said, participants provided examples of how inter- and intraregional movement of clinical staff helped to retain the necessary skill sets within Western Australia. The extent to which clinical skills acquired through ETS involvement in ED consultations benefited the overall capability of the regional clinical workforce in WA fell outside the scope of this paper; however, this is an area of significant importance and relevance to the effectiveness of the ETS.

The pivotal but nonmodifiable consideration in clinical decision-making was the physical distance from definitive treatment locations and presenting hospitals. Local hospital resources, capacity and capability were beyond the scope of this study, but it was acknowledged that these had limited modifiability, at least in the short term.

This paper did not explore the impact of the ETS on the quality and standard of care in larger regional EDs, and the dominant view that EM specialists are risk managers is of significance. Any manifested risk aversion is likely related to patient safety considerations given that rural and remote contextual factors are nonmodifiable or have limited modifiability. Despite the dominant view that the ETS added value to presenting hospital clinical decisions, there were also concerns that pre-existing processes and capacity had been disrupted by the care processes now available through the ETS. The change in the scope of practice in presenting hospitals, particularly regarding nursing staff, because of ETS involvement is a subject for further detailed exploration to reach an optimal balance.

Novel Contributions of This Study

Prior work has paid considerable attention to advocating for telehealth policy reform [30] and identifying the barriers and enablers of telehealth implementation in rural and remote EDs. This is the first study to systematically detail the complex interrelationships between technological and human interfaces. The barriers to quality and safety of emergency telehealth consultations and mitigation strategies used by the WACHS ETS clinicians when collaborating with rural and remote EDs were detailed. These perceived clinical risks and risk mitigation strategies added new insights into the contextual factors influencing the effective implementation of rural and remote ETS.

Findings around technological challenges were consistent with the findings of previous studies that pointed to variations in usability, depending on the setting and situation. Unlike previous studies that compared the approach of different types or combinations of telehealth technologies (eg, videoconferencing, store and forward, and telemonitoring) [21], this study embraced the use of available technology as a norm in rural and remote ETS practices in Western Australia. The focus on the strategies used in WACHS ETS delivery to overcome technological barriers in this study makes a novel contribution to advancing our knowledge on the modifiability of the risks associated with rural and remote virtual care and helps to inform future quantitative analysis and interpretation of findings on the effectiveness of rural and remote ETS.

Limitations and Future Research

The findings reported in this paper are subject to several limitations. Clinical nurse coordinators are the first contact points for accessing the ETS, and this triaging (referred to as ETS prioritization to distinguish from the Australasian Triage Scoring system) plays a pivotal role in case review order and selection. The ETS prioritization process ensured that the presenting hospital and community context were incorporated into timely and appropriate care from the first point of ETS contact. This paper focused on regional context considerations influencing clinical decisions around the disposition of ED patients during a virtual consultation and did not examine regional context factors already incorporated during the process of ETS prioritization. Future reports on ETS prioritization would be an essential addition to the body of knowledge on emergency telehealth implementation in the rural and remote context.

Telehealth, especially emergency telehealth, is a relatively new area in clinical practice with evolving clinical governance standards. Medical indemnity issues were discussed during the interviews but did not emerge as a dominant theme in the data analysis. WACHS ETS physicians are particularly focused on patient safety. However, based on available data, it is difficult to conclude the extent to which medicolegal considerations are at play in the response to the risks associated with the ETS practice.

The financial viability or cost-effectiveness of virtual care is best explored quantitatively. It is worth noting that participants were value conscious in evaluating disposition options. However, they were not able to comment on the financial viability of the service as it was not within the role of most of these clinician participants. The implementation factors identified inevitably influence the effectiveness of telehealth in rural and remote EDs but are insufficient to determine the effectiveness of the ETS. This study represents a step toward determining the effectiveness of ETS delivery and is an area for future research.

The recruitment of participants may have been subject to social desirability bias, and some nonresponders may have had an alternative opinion about the service, which they felt did not align with the health service's official position. The presenting health condition is one of the central tenets in this research, although the data collection did not focus on specific health conditions. Our intention in this paper was to capture the overarching themes of clinical decision-making and how they affect the safe and effective delivery of rural and remote ETS to improve patient health outcomes. There were no comparisons between the treatments of urgent and nonurgent presentations because formative analysis showed the tendency of using the ETS for more urgent cases. This focus was due to the higher proportion of FACEMs among the study participants, who were more likely to encounter urgent cases.

Conclusion

Telehealth has the potential to improve the processes of clinical decision-making in rural and remote EDs. The participants, all

of whom participated in delivering the ETS, perceived improved accuracy, confidence, and speed of clinical decisions. The reach and effectiveness of the ETS in addressing the identified health needs remain to be tested.

Patient safety was central to all clinical decisions, which were grounded in the capability of presenting hospitals and the local community context. The risks of delay in definitive care for those patients requiring transfer for ongoing management associated with the geographical distance to the hospital for definitive treatment and the presenting hospital capability were not amenable to change. The importance of social and geographical contexts for appropriate telehealth delivery was highlighted in this study.

Technological barriers to the safe and effective delivery of the ETS were amenable to improvisation and improvements to infrastructure, but overall, participants were aware that they could not reach the same reliability as that in face-to-face consultation for all presenting conditions. Although the ETS was an acceptable and feasible option for ED consultation, the maintenance and sustainability of the ETS as part of routine practice and policy are dependent on presenting conditions.

Investment into rural and remote telehealth has the potential to support regionally based clinicians both directly in patient care and indirectly by promoting their (health workforce) longevity in regional areas through the ongoing development of the virtual care network. The effect of the ETS on the scope of the regional ED practice and the building of clinical skills within the region warrants further in-depth exploration including its potential implications for the overall effectiveness and cost-effectiveness of telehealth services in rural and remote EDs.

Acknowledgments

This research was initiated by the WA Country Health Services (WACHS) and Curtin University, and received financial support from both institutions as well as the Digital Health Collaborative Research Centre (DHCRC) in the form of PhD scholarships. The WACHS also contributed in-kind support in the form of clinician time, industry supervision, and administrative and research infrastructure support. This research received further in-kind contributions from the School of Population Health at Curtin University, WA Centre for Rural Health at the University of Western Australia, WACHS Command Centre, and WACHS Country Health Innovation Project Support Office in the form of academic supervision, industry mentorship, and knowledge exchange.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide for Emergency Telehealth Service clinicians.
[DOCX File , 18 KB - humanfactors_v12i1e58851_app1.docx]

Multimedia Appendix 2
The professional designations and perspectives of study participants.
[DOCX File , 17 KB - humanfactors_v12i1e58851_app2.docx]

References

1. Sikka N, Gross H, Joshi AU, Shaheen E, Baker MJ, Ash A, et al. Defining emergency telehealth. J Telemed Telecare 2019 Dec 11;27(8):527-530. [doi: 10.1177/1357633x19891653]
2. Interim Position Statement: Teleheatlh in Emergency Medicine (S843). Australasian College for Emergency Medicine. URL: https://www.mdanational.com.au/-/media/files/pdfs-linked-to-library-items/s843_interim_statement_telehealth.pdf [accessed 2025-01-30]
3. Public consultation - Guidelines: Telehealth consultations with patients. Medical Board of Australia. URL: https://www.medicalboard.gov.au/News/Past-Consultations/2023.aspx [accessed 2023-07-24]
4. WA Country Health Service. Emergency Department Data Collection. Perth, Western Australia: WA Country Health Service; 2025.
5. Emergency Telehealth Service. Government of Western Australia. URL: https://www.wacountry.health.wa.gov.au/Our-patients/Telehealth/Emergency-Telehealth-Service [accessed 2025-01-30]
6. Command Centre. Government of Western Australia. URL: https://www.wacountry.health.wa.gov.au/Our-services/Command-Centre [accessed 2025-01-30]
7. WA Country Health Service. Emergency Telehealth Collection. Perth, Western Australia: WA Country Health Service; 2023.
8. Kolhatkar A, Keesey A, Bluman B, Lynn B, Wilkinson T. Understanding how emergency medicine physicians survive and thrive in rural practice: a theoretical model. Rural and Remote Health 2017;17(4):1-14. [doi: 10.22605/rrh4285]

9. Setrinen Hansen N, Mikkelsen S, Bruun H, From E, Milling L. Physicians' experiences working in emergency medicine in a rural area in Northern Sweden: a qualitative study. Rural and Remote Health 2021 Dec;21(4):1-11. [doi: 10.22605/rrh6672]
10. du Toit M, Malau-Aduli B, Vangaveti V, Sabesan S, Ray RA. Use of telehealth in the management of non-critical emergencies in rural or remote emergency departments: A systematic review. J Telemed Telecare 2017 Oct 05;25(1):3-16. [doi: 10.1177/1357633x17734239]
11. Brainard JS, Ford JA, Steel N, Jones AP. A systematic review of health service interventions to reduce use of unplanned health care in rural areas. J Eval Clin Pract 2016 Apr 28;22(2):145-155. [doi: 10.1111/jep.12470] [Medline: 26507368]
12. Culmer N, Smith T, Stager C, Meyer H, Quick S, Grimm K. Evaluation of the triple aim of medicine in prehospital telemedicine: A systematic literature review. J Telemed Telecare 2019 Jun 25;26(10):571-580. [doi: 10.1177/1357633x19853461]
13. Sharifi Kia A, Rafizadeh M, Shahmoradi L. Telemedicine in the emergency department: an overview of systematic reviews. Z Gesundh Wiss 2023;31:1193-1207 [FREE Full text] [doi: 10.1007/s10389-021-01684-x] [Medline: 35103232]
14. Kelton DK, Szulewski A, Howes D. Real-time video telemedicine applications in the emergency department: a scoping review of literature. Canadian Journal of Emergency Medicine 2018 Nov 22;20(6):920-928. [doi: 10.1017/cem.2017.382] [Medline: 28829008]
15. Nadar M, Jouvet P, Tucci M, Toledano B, Sicotte C. Impact of synchronous telemedicine models on clinical outcomes in pediatric acute care settings: a systematic review. Pediatric Critical Care Medicine 2018 Dec;19(12):e662-e671. [doi: 10.1097/PCC.0000000000001733] [Medline: 30234678]
16. Winburn AS, Brixey JJ, Langabeer J, Champagne-Langabeer T. A systematic review of prehospital telehealth utilization. Journal of Telemedicine and Telecare 2017 Jun 22;24(7):473-481. [doi: 10.1177/1357633x17713140]
17. Lazarus G, Permana AP, Nugroho SW, Audrey J, Wijaya DN, Widyahening IS. Telestroke strategies to enhance acute stroke management in rural settings: A systematic review and meta-analysis. Brain and Behaviour 2020 Oct 18;10(10):e01787 [FREE Full text] [doi: 10.1002/brb3.1787] [Medline: 32812380]
18. Ward MM, Jaana M, Natafgi N. Systematic review of telemedicine applications in emergency rooms. International Journal of Medical Informatics 2015 Sep;84(9):601-616. [doi: 10.1016/j.ijmedinf.2015.05.009] [Medline: 26072326]
19. Boggan JC, Shoup JP, Whited JD, Van Voorhees E, Gordon AM, Rushton S, et al. Effectiveness of acute care remote triage systems: a systematic review. J Gen Intern Med 2020 Jul 02;35(7):2136-2145 [FREE Full text] [doi: 10.1007/s11606-019-05585-4] [Medline: 31898116]
20. Eder PA, Reime B, Wurmb T, Kippnich U, Shammas L, Rashid A. Prehospital telemedical emergency management of severely injured trauma patients. Methods of Information in Medicine 2018 Nov;57(5-06):231-242. [doi: 10.1055/s-0039-1681089] [Medline: 30875702]
21. Rogers H, Madathil KC, Agnisarman S, Narasimha S, Ashok A, Nair A, et al. A systematic review of the implementation challenges of telemedicine systems in ambulances. Telemedicine Journal and E-Health 2017 Sep;23(9):707-717. [doi: 10.1089/tmj.2016.0248] [Medline: 28294704]
22. Salmoiraghi A, Hussain S. A systematic review of the use of telepsychiatry in acute settings. Journal of Psychiatric Practice 2015 Sep;21(5):389-393. [doi: 10.1097/PRA.0000000000000103] [Medline: 26348806]
23. Kimmel HJ, Brice YN, Trikalinos TA, Sarkar IN, Ranney ML. Real-time emergency department electronic notifications regarding high-risk patients: a systematic review. Telemedicine Journal and E Health 2019 Jul;25(7):604-618 [FREE Full text] [doi: 10.1089/tmj.2018.0117] [Medline: 30129886]
24. Marsh-Feiley G, Eadie L, Wilson P. Telesonography in emergency medicine: A systematic review. PLoS One 2018 May 3;13(5):e0194840 [FREE Full text] [doi: 10.1371/journal.pone.0194840] [Medline: 29723198]
25. Gattu R, Teshome G, Lichenstein R. Telemedicine applications for the pediatric emergency medicine. Pediatric Emergency Care 2016;32(2):123-130. [doi: 10.1097/pec.0000000000000712]
26. Pak D, Pak K. Effectiveness of telemedicine in acute / emergency care settings versus face to face patient care: a systematic literature review. Journal of the International Society for Telemedicine and eHealth 2016;3:e14-1-e14-13 [FREE Full text]
27. Is Prehospital Emergency Telemedicine Implementation Feasible In Non‐Traditional EMS Settings: A Systematic Literature Review. The University of Arizona. 2017. URL: https://repository.arizona.edu/handle/10150/623626 [accessed 2025-01-30]
28. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: SAGE Publications, Inc; 2014.
29. Strauss A, Corbin JM. Grounded Theory in Practice. Thousand Oaks, CA: SAGE Publications, Inc; 1999.
30. Rangachari P, Mushiana SS, Herbert K. A scoping review of applications of the Consolidated Framework for Implementation Research (CFIR) to telehealth service implementation initiatives. BMC Health Serv Res 2022 Nov 30;22(1):1450 [FREE Full text] [doi: 10.1186/s12913-022-08871-w] [Medline: 36447279]

Abbreviations

ACEM: Australasian College of Emergency Medicine
ANZFTCOP: Australian New Zealand FACEM Telemedicine Community of Practice

ED: emergency department
EM: emergency medicine
ETS: Emergency Telehealth Service
FACEM: Fellow of Australasian College of Emergency Medicine
GP: general practitioner
WACHS: WA Country Health Service

Edited by A Kushniruk, D Kaufman; submitted 26.03.24; peer-reviewed by J Job, GA Kingston, N Harvey; comments to author 07.07.24; revised version received 01.12.24; accepted 23.01.25; published 15.04.25.

Please cite as:
Tsou C, Yeung J, Goode M, Mcdonnell J, Williams A, Andrew SC, Tetlow J, Jamieson A, Hendrie D, Reid C, Thompson S
Perceived Risks, Mitigation Strategies, and Modifiability of Telehealth in Rural and Remote Emergency Departments: Qualitative Exploration Study
JMIR Hum Factors 2025;12:e58851
URL: https://humanfactors.jmir.org/2025/1/e58851
doi:10.2196/58851
PMID:40232817

Original Paper

Experiences of Wheelchair Users With Spinal Cord Injury With Self-Tracking and Commercial Self-Tracking Technology ("In Our World, Calories Are Very Important"): Qualitative Interview Study

Vasiliki Mylonopoulou[1], PhD; Katerina Cerna[2], PhD; Alexandra Weilenmann[1], Prof Dr; Mattias Rost[1], PhD; Tobias Holmlund[3], PhD

[1]Applied IT, University of Gothenburg, Gothenburg, Sweden

[2]School of Information Technology, Halmstad University, Halmstad, Sweden

[3]Neurobiology, Care Sciences and Society, Karolinska Institute, Stockholm, Sweden

Corresponding Author:
Vasiliki Mylonopoulou, PhD
Applied IT
University of Gothenburg
Forskningsgången 6
Gothenburg, 417 56
Sweden
Phone: 46 0702517497
Email: vasiliki.mylonopoulou@gu.se

Abstract

Background: Commercial wearable and mobile wellness apps and devices have become increasingly affordable and ubiquitous. One of their aims is to assist the individual wearing them in adopting a healthier lifestyle through tracking and visualizing their data. Some of these devices and apps have a wheelchair mode that indicates that they are designed for different types of bodies (eg, wheelchair users with spinal cord injury [SCI]). However, research focuses mainly on designing and developing new condition-specific self-tracking technology, whereas the experiences of wheelchair users with SCI using self-tracking technology remain underexplored.

Objective: The objectives of this study were to (1) provide a comprehensive overview of the literature in the field of self-tracking technology and wheelchair users (as a basis for the study), (2) present the self-tracking needs of wheelchair users with SCI, and (3) present their experiences and use of commercial self-tracking technology.

Methods: We conducted semistructured interviews with wheelchair users with SCI to understand their experiences with self-tracking and self-tracking technologies, their self-tracking needs, and how they changed before and after the injury. The interviews were thematically analyzed using an inductive approach.

Results: Our findings comprised three themes: (1) being a wheelchair user with SCI, (2) reasons for self-tracking, and (3) experiences with self-tracking technologies and tools. The last theme comprised 3 subthemes: self-tracking technology use, trust in self-tracking technology, and calorie tracking.

Conclusions: In the Discussion section, we present how our findings relate to the literature and discuss the lack of trust in commercial self-tracking technologies regarding calorie tracking, as well as the role of wheelchair users with SCI in the design of commercial self-tracking technology.

(JMIR Hum Factors 2025;12:e65207) doi:10.2196/65207

KEYWORDS

wheelchair; spinal cord injury; tracking; self-tracking; wellness technology; calories; health inequalities; inclusive design in mobile health; design; lifestyle app; artificial intelligence; AI

Introduction

Background

Every year, up to 500,000 people experience spinal cord injury (SCI) and, as a result, become wheelchair users [1]. This population must adjust to their new body, learn to use a wheelchair, and sustain an active lifestyle. An active lifestyle is vital for their independent living and quality of life. If they are not active, their shoulders can be injured because they are overused in their daily life without being trained for specific activities (eg, getting in and out of the wheelchair). Another risk is that they can gain weight, which requires them to change to a wider wheelchair, thus rendering more places inaccessible.

As commercial self-tracking technology (eg, smartwatches) becomes increasingly common, wheelchair users can access and benefit from its advancements. Some of this technology (eg, Apple Watch and Runkeeper) has already implemented accessibility settings for wheelchair users. Up to now, research on self-tracking technology has focused on (1) chair-ables, artifacts placed on or in the wheelchair to promote a healthier lifestyle [2-4]; and (2) studies on wheelchair users' experiences with self-tracking technology that was provided or introduced to them [5-7]. However, the everyday lived experience of self-tracking and related technology (ie, self-tracking technology) of wheelchair users with SCI is underresearched. This can result in the exclusion and underrepresentation of this specific group in the research and development of commercial wearables. Excluding part of the population as "special users" is an ethical and social issue [8,9]. In addition, more commercial technologies are anticipated to become accessible to a broader audience once the European Accessibility Act (EAA) is applied. This act focuses on digital products and services, underlining the importance of mainstream technology being accessible to people with disabilities and older adults [10].

The objective of this study is to build on previous research on self-tracking and self-tracking technology among wheelchair users with SCI [5-7,11] to include lived experience of self-tracking and self-tracking technology. Through an interview study, we aimed to answer the following question: How do wheelchair users with SCI experience self-tracking and commercial self-tracking technology in their everyday lives?

This section first provides an overview of SCI and the impact of the injury on people's lives. This is followed by a short overview of existing research on wheelchair users and self-tracking technology for health and well-being. Finally, we position our research in the context of contemporary society and technological research focusing on the role of wheelchair users with SCI.

SCI Overview

Worldwide, approximately 90 million people live with SCI, many of whom are either underage or part of the working population [1,12]. Each year, between 250,000 and 500,000 people sustain an injury to their spinal cord [1]. The average lifetime costs of an SCI starting at the age of 25 years range from €0.45 million to €2.1 million (US $0.49 million-$2.3 million), which exceeds those related to dementia, multiple sclerosis, and cerebral palsy [1]. Poor quality of life for people with SCI, their families, and caregivers—in combination with the socioeconomic burdens caused by the injuries—make it imperative to find solutions for this condition.

A motor-complete SCI leads to complete muscle mass loss in the legs and, thus, life as a wheelchair user. For a walking person, the muscles in the legs are the main consumers of energy both in an active state (eg, when the person exercises or moves) and in a rest state (eg, sleeping or sitting) [13,14]. In addition, an injury above thoracic level 6 provokes an altered (physiological) autonomic response during both rest and activity [15]. This results in an altered metabolic response, leading to reduced energy burn in both states [13-15]. Body weight balance is partly explained as caloric balance based on calorie intake and output [16-18]. After an SCI, the total energy burn (ie, calorie burn) is much lower than before. Generally, after the initial acute phase, people with SCI lose weight due to the loss of muscle mass and the healing process. During rehabilitation, most people get used to their light body weight and narrow wheelchair, facilitating an active and independent daily life. Throughout the rehabilitation phase, body weight tends to increase (approximately ≥2 kg during the first year) [19]. This relates to the lack of compensatory caloric intake due to the lower metabolic demands (total daily energy burn), which leads to energy imbalance with higher calorie intake than energy burn [19]. Increased body weight can have many consequences, such as higher stress on shoulder muscles due to extra weight being carried each time the person transfers to and from their wheelchair [20]. Moreover, it becomes more difficult to manage daily activities such as self-care. The resulting shoulder pain and reduced independence are often devastating [20]. Thus, calories are extremely important for people with SCI. Maintaining an appropriate calorie balance becomes much more achievable when energy burn and calorie intake can be monitored.

Wheelchair Users With SCI and Self-Tracking Technology for Health and Well-Being

Overview

Self-tracking is a voluntary, reflexive practice in which people collect data about themselves to enhance self-awareness and understanding by observing, noting, and interpreting various aspects of their lives [21]. Self-tracking technology is any technology that supports people in self-tracking practice. This technology can include apps—where the user registers information about their health and well-being—smartwatches or other wearable technology (wearables), or sensors embedded in assistive technology such as a wheelchair (chair-ables). Self-tracking technology is not assistive technology as it does not aim to improve the functional ability of people with disabilities or enable and enhance their participation and inclusion in different domains of life [22]. However, self-tracking technologies can offer inclusivity through adjusting functionality to include people with disabilities.

In this study, we focused on the experiences of people with SCI with mainstream technology as we aimed to contribute to an inclusive design. However, we recognize the importance of

understanding and learning from assistive technology for our target demographic. This understanding could direct inclusive design research and practice as it provides insights into adjusting or expanding the mainstream technology to be more inclusive.

Assistive Technology

Chair-ables have been researched in terms of tracking specific behaviors and self-care–related injuries. Let us take the case of pressure releases, which are vital for wheelchair users in avoiding pressure ulcers potentially contributing to premature death [23]. A chair-able can support wheelchair users with pressure release exercises. Sensors can be placed on the wheelchair to measure in-seat movement, classify the activity from the sensors into weight shifts, and inform the user with their data as well as sending push notifications upon achieving goals related to pressure releases [2]. In 2021, Ahad et al [4] developed and evaluated the reliability of an algorithm and a chair-able that could support people on wheelchairs in doing their pressure release exercises. The year after, design considerations for supporting electric wheelchair users with SCI on their pressure releases were published [24]. The authors [24] implemented context-aware and unobtrusive reminders delivered by the wheelchair for incomplete or incorrect pressure release exercises. The interaction with the wheelchair in this context [24] was preferable to an interaction with an app. Similarly, but in the context of self-tracking apps, Amann et al [25,26] used participatory design to design, develop, and evaluate an evidence-based app that uses a smart camera to prevent pressure injuries.

As for self-care, Büyüktür et al [3] presented requirements for semiautomated tracking to support people with SCI in their self-care by interviewing patients and health care professionals. Their main findings were that the patients' routines change from the clinic to the home environment and each patient decides to focus on what they value the most (eg, sleeping instead of waking up to change position during their sleep). Each patient and their informal caregivers try different routines in the home environment until they find one through which they can track the behavior and with which they can comply. Thus, Büyüktür et al [3] recommended a highly tailored and flexible system that can follow the patient's routines as they change and give them feedback. In general, a systematic literature review [11] showed that there is research on self-management of SCI in the home environment and clinic as well as the promotion of physical activity and a healthy lifestyle (eg, wellness and fitness). For example, one of the articles reviewed [27] presented that an intervention that tracked the activity of wheelchair users with SCI and motivated them through just-in-time messaging had the potential to improve physical activity for wheelchair users with SCI. Similarly, an interview study [28] explored the barriers that people with SCI face to staying physically active: (1) lack of tailored physical activity forums with wheelchair users with SCI and (2) lack of personalized fitness-tracking technology tailored to wheelchair-based activities.

Mo et al [29] explored the information needs of wheelchair users, and they identified a lack of tailored, accurate, and affordable tracking for wheelchair users both commercially and academically. Even though they argued for inclusive technology, the start of filling this gap could be in the assistive technology literature. Li et al [30] introduced WheelPoser, a system that tracks and identifies wheelchair users' on-the-move activity. They argued that they could do that by using 4 small sensors. Their dataset, code, and models are available for everyone to use, which allows mainstream technology to become more inclusive by using the same system as an add-on to their trackers or only the dataset to train their own systems. Huang et al [31], recognizing the lack of data on wheelchair users for training artificial intelligence algorithms on pose estimation, developed and evaluated a dataset to fill that gap. Their dataset is publicly available as well. Another article published in a disability forum explored the relationship between the successful completion of activity guidelines for people with SCI and the fitness or health status of the person with SCI [32]. This paper referred to activity guidelines [33,34] for people with SCI and an understanding of how mainstream technology can tailor its features (eg, daily goals) to a specific SCI group.

Inclusive Technology

Research underlines the need for inclusive and accurate commercial technology for wheelchair users through studies [35] and literature reviews [36,37]. The participants in the study by Li et al [35] showed interest in tracking their activity, but they did not trust the accuracy of contemporary commercial technology, understanding that those using powered wheelchairs could track some data on their activity that were relevant to them, such as distance. A literature review explored how academic publications in the personal informatics field address the information needs of wheelchair users [36]; it consolidated recommendations to tackle the design challenges that researchers and practitioners may face when attempting to design for including wheelchair users' activity tracking and information needs. In 2019, Moon et al [37] conducted a literature review of studies on the inclusivity of contemporary digital technology. They advocated for inclusive and universal design to become an integral part of commercial digital technology's development process and consolidated methods that can support the process.

Another systematic literature review [11] published in 2023 examined the use of mobile health in supporting people with SCI to maintain or improve their health. It showed that research lacked results regarding the role of commercial self-tracking technology for health and wellness to support people with SCI. The authors [11] attributed this gap in research to the fact that commercial self-tracking devices are inaccurate in their calculations for wheelchair users. However, in 2021, a study [38] evaluated the accuracy of the Apple Watch Series 1, showing that the watch was accurate for high-frequency strokes on a wheelchair (eg, a treadmill) but not for low-frequency strokes (eg, walking). Even though the research was published recently, the results of this research may not correspond to the current version of the Apple Watch (Series 7).

Related to commercial trackers and their use by of wheelchair users with SCI or their preferences, there is research limited to commercial technology given to them for a specific amount of time. For example, Ungurean and Vatavu [7] published a study in 2022 exploring the needs of wheelchair users related to activity trackers regarding different self-tracking technologies

(eg, rings, watches, armbands, and chair-ables). The researchers [7] interviewed participants after letting them watch videos of 2 different self-tracking technologies (eg, ring and armband) to understand their perceptions, preferences, and willingness to use wearables. Some of their participants were already wearable technology users, and the researchers mentioned that this could have supported them in answering the interview questions [7]. However, the study's results focused on the videos that the participants watched.

Malu and Findlater [5] based their research on the work by Carrington et al [6]. Carrington et al [6] tracked variables of wheelchair basketball players and presented the variables to the athletes to understand their needs and potential data use. In addition, Malu and Findlater [5] included wheelchair users who were not necessarily athletes. Their participants were interviewed, participated in design workshops, and tried 2 commercial technologies. Specifically, they used 3 fitness trackers for 30 minutes before they compared them and participated in a design session for designing a fitness tracker. Finally, both studies agreed that the values that people in wheelchairs may want to see are energy burned (calories), distance, GPS data, and pulse. The participants in the study by Malu and Findlater [5] added that calorie intake is equally important to energy burned. Finally, Malu and Findlater [5], unlike Carrington et al [6], showed that commercial self-tracking devices using GPS and comparing mobility in steps between different dates were acceptable and of value to their participants.

In 2019, Helle and Rosenbeck Gøegb [39] conducted a focus group interview study with 7 wheelchair users who were members of a basketball team. The study aimed to explore wheelchair users' experiences with current self-tracking devices and future requirements. Their key results showed that their participants only found the distance and time parameters useful and perceived calorie consumption, pulse, steps, and training intensity as inaccurate [39]. Their study participants commented that "some activity trackers provide reminders that assume that the user can walk" [39]. The authors concluded that accurate activity trackers designed for wheelchair users are needed. It is not obvious from this publication whether the participants were daily users of the technology or whether the technology was given to them for study purposes, and we do not know whether they were professional athletes.

To summarize, research on self-tracking and self-tracking technology and its use by wheelchair users with SCI is primarily focused on either building technology for them or gathering requirements and their opinions on contemporary commercial self-tracking technology, with which they may have limited experience [5-7,11,25,39]. Our research enriches this body of knowledge [5-7,11,25,39] with lived experience of people who track or tracked their activity or use or used commercial self-tracking technology. This study also increases the body of knowledge of everyday lived experiences with self-tracking devices and self-tracking for an often underrepresented community in technological research—wheelchair users with SCI [40].

More Reasons for Inclusion in the Design of Commercial Technology

People with disabilities are often considered users of "special" products designed explicitly for them instead of active participants in the mainstream market [41]. "Special" targeted products would have a smaller share in the market, and they may be more expensive and, therefore, inaccessible for many people with disabilities as people with disabilities—at least in the European Union (EU) [42]—tend to be at a higher risk of financial challenges. According to Eskyte [41], dividing the market into "average" and "vulnerable" users is an obstacle to incorporating accessibility requirements into general consumer product development. In 2025, all digital services and products in the EU market need to be accessible according to the EAA [10], which will urge practitioners to design digital products and services with a diverse audience in mind.

Publications in technological fields support that building things for people with disabilities often means building better things for everyone [9], and a systematic literature review [11] shows that research on technology for people with SCI tends toward user-centered design. Even though user-centered design focuses on the user by collecting information about them, building empathy—often by simulating their health conditions [43]—and potentially involving them in key stages of the design and development process, it does not seem to be enough. Bennett and Rosner [44] argued for a more participatory way of designing; instead of designing for the person who is different from the designers (eg, wheelchair users), we should design *with* them as part of the design team. Their research indicates that, by trying to simulate their life experiences, we do not treat them as equals but as something different from us, different from the norm [44]. While the solution to the limitations of current empathy-building methods seems to be the participation of the specific user group in the design process, participation does not guarantee technology acceptance [26] and may be hard to achieve [40]. For example, in the case of SCI, Kabir et al [40] support that, due to the impact of the condition on the user, for example, speaking or fatigue issues may have been excluded from the research because the collection methods of the designers and researchers are inaccessible. Kabir et al [40] describe that the methods need to be adjusted based on each participant's abilities, something that some publication forums and reviewers may consider a methodological weakness of a study.

This paper does not argue about replacing the involvement of people with SCI in research or user research. However, it gives an idea of how current people with SCI experience commercial self-tracking technology and describes some of their needs. Even though researchers and practitioners will need to involve people with SCI when designing accessible products, this study can act as a starting point to understand some people's experiences with SCI, their self-tracking habits, and their use of self-tracking technology.

Methods

We will first present the research context and then how the data collection and analysis were conducted.

Context of the Study

This research was conducted as part of a long-lasting cooperation among researchers in the human-computer interaction field; a researcher in physiotherapy focusing on wheelchair users with SCI; and a center for SCI in Gothenburg, Sweden. This cooperation aims to explore the use of affordable commercial wearables and build an application that uses the wearables' data to accurately calculate the collected variables specifically for wheelchair users with SCI. The project is based on research exploring the energy outtake for wheelchair users with SCI on specific exercises [21-24]. Within the scope of this project, we conducted interviews to understand *how wheelchair users with SCI experience self-tracking and the use of commercial self-tracking technology in their everyday lives*.

Ethical Considerations

At the start of the cooperation, the planned research was reviewed and evaluated for its ethical integrity following the process of the Department of Applied IT at the University of Gothenburg, Sweden. This qualitative study interviewed people with SCI on their tracking habits and use of commercial self-tracking technology. This type of research does not fall under any of the categories described by the Riksdag (the supreme decision-making body of the Kingdom of Sweden) as research requiring ethical clearance as no sensitive personal data or other personal data were collected or published [45]. The study did not have access to or collect interviewees' personal information prior to, during, or after interviews. The participants were informed about the study through an informed consent form and could ask questions before taking part. The information sheet also included details on how they could withdraw. The data of the participants were handled according to the Swedish implementation of the European General Data Protection Regulation. Participants received no compensation.

Recruitment, Data Collection, and Data Analysis

We conducted 9 semistructured interviews [46,47] with wheelchair users with SCI. Knowing from the literature [40] that we may face challenges in recruiting participants for interviews, we used nonprobability convenience sampling using the network of the SCI center in Gothenburg, Sweden, and personal connections. Eligible for participation were people aged >18 years who had SCI for more than a year and who used a wheelchair to move. If participants did not fulfill all the aforementioned criteria, they were excluded. The aim was to have a deep understanding of the experiences of the people who contacted us rather than having a certain quantity of data.

Due to the geographical distribution of the participants and to accommodate the interviews at any time the participants wanted (even if it was on short notice), we conducted the interviews and recorded them over Zoom video call (Zoom Video Communications). In that way, the participants could also show us how they tracked themselves if needed. The interview guide had three groups of questions: (1) introductory and background questions, aiming to get to know the interviewees and their habits concerning daily life and exercise activity, including mobility; (2) questions about experiences of self-tracking and self-tracking technologies before and after their injury, aiming to understand the relationship of the interviewee with self-tracking and self-tracking technologies before and after their injury; and (3) closure of the interview questions, aiming to allow the interviewees to add anything they wanted us to know and understand their interest in the results of the study. We did not collect any actual measurements of their physical activity, nor did we try to evaluate their physical activity based on any tool as the focus of this study was to understand their experiences with self-tracking and self-tracking technologies.

The average duration of the interviews was 45 (SD 12.9) minutes, with the shortest being 30 minutes long and the longest being 69 minutes long (390 min; 6.5 h of video recordings in total). All participants were Swedish; 89% (8/9) of the interviews were conducted in English, and 11% (1/9) were conducted in Swedish. The participants interviewed in English could communicate clearly and fluently in English. The interviews were transcribed, and the first author analyzed them through inductive thematic analysis [48] following the steps by Braun and Clarke [49]. The first author familiarized themselves with the data and found codes and initial themes (steps 1-3). All authors reviewed, defined, and named the themes (steps 4-5). In our qualitative and explorative study, we focused on providing a rich description of our participants' experiences with self-tracking and self-tracking technology rather than the frequency with which they talked about each theme. The quotes used were modified to remove repetitions and for grammatical correctness. The quotes taken from the Swedish-language interview were translated into English.

Methodological Limitations

To invite the participants, we used the connections mentioned previously. In the invitation, we informed participants that it would be an online interview; thus, participants who had difficulty speaking due to their injury or other reasons may have been excluded [40]. We expected a small sample who would have the time and energy to participate. To counteract this, we invited people with SCI to participate regardless of whether they were current or past users of self-tracking technology as long as they had tracked something related to their injury or physical activity at some point. We focused on analyzing and obtaining a deeper understanding of the experiences of the 9 people rather than on collecting more data.

Results

Overview

A total of 22% (2/9) of the participants were female, reflecting the gender ratio of SCI at 20% female [50]. In total, 22% (2/9) of the participants had never used self-tracking technology, but they had all tracked some variables. Apart from participant P5, no one was currently an athlete. All the participants (9/9, 100%) were well past their first year of living with SCI, characterized as phase 3 or the chronic phase [40]. No participant had any serious issues with breathing or speaking during the interview regardless of their level of injury. More details about each participant's sex, age, years with SCI, and use of self-tracking technology, as well as the code used to refer to them in the following sections, can be found in Table 1.

Table 1. Demographics of the participants, including their use of self-tracking technology.

Participant code	Sex	Age group (y)	Years with SCI[a]	Use of self-tracking technology
P1	Male	50s	35	Yes
P2	Female	40s	31	Yes
P3	Male	30s	5	Yes
P4	Male	≥65	29	No
P5	Male	30s	12	Yes
P6	Male	40s	13	Yes
P7	Female	≥65	40	No
P8	Male	60s	6	Yes
P9	Male	50s	—[b]	Yes

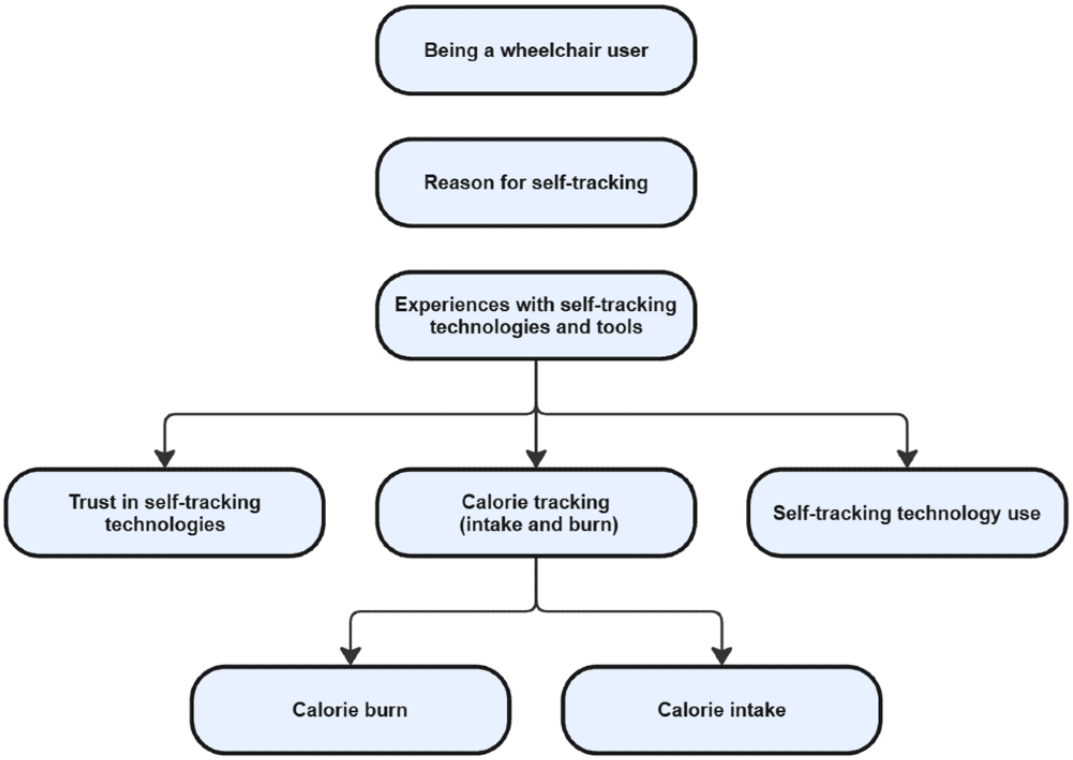
[a]SCI: spinal cord injury.

[b]Not applicable.

In total, 78% (7/9) of the participants had extensively experienced self-tracking technology; however, this high level of engagement with technology should be considered within the context of the study. People living in Sweden use IT daily for mundane things (eg, mobile payment is common; 92% of the population) [51]. In addition, Sweden is among the 7 countries in the EU with the lowest gap in employment between people with and without disabilities, which may have allowed our participants to have enough income [52] to purchase self-tracking technologies.

The thematic analysis resulted in 3 themes, one of which had 3 subthemes. The first theme, *being a wheelchair user with SCI*, included codes relevant to the participants' descriptions of their activity levels, themselves, and their lives. The second theme, *reasons for self-tracking*, included codes related to the participants' self-tracking habits. The third theme, *experiences with self-tracking technologies and tools*, included subthemes related to self-tracking technology use, calorie tracking (intake and burn), and trust in self-tracking technologies. Figure 1 provides a more illustrative understanding of how the themes are structured. The figure is designed to be read from top to bottom and left to right. A tabular form of the figure can be found in Multimedia Appendix 1.

Figure 1. The final themes that resulted from the thematic analysis in a diagrammatic representation. Multimedia Appendix 1 shows the representation of the themes in a tabular form with a summary of the main findings.



Being a Wheelchair User With SCI

This presentation of the participants aims to increase our empathy and understanding of how they experience their lives as wheelchair users and as former walking persons. Wheelchair users with SCI experience life first as walking people and then as wheelchair users. This means that they must adjust to their new body and develop new skills and often see themselves and their life divided into 2 periods—before and after the injury. For example, P8, who had a physical job and changed to a desk job after becoming a wheelchair user, mentioned that "before the injury, I did not track [my activity]. You could see I was physically in a good place. I was very masculine. After the injury, it is another story...you need to train, of course, a lot of rehabilitation." The various SCIs can influence people's bodies and lives differently, widening the diversity of their bodies.

Our participants dealt with the change in their skills in five ways: (1) adjusting their previous knowledge to the new situation, (2) giving up old habits, (3) being challenged to keep old habits, (4) reflecting on the behavior of their body, and (5) acquiring new skills. For example, regarding item 1, P3 adjusted his previous knowledge to the new situation in the following way:

> *I knew what because of my experience before the injury. I knew how to train, [but] I had to adapt to the situation I got into. I could not do everything in the same way, but I got some tips from the rehabilitation clinic, and I tested them myself.*

Others had advanced skills as non–wheelchair users that they lost when they became wheelchair users. For example, regarding item 2, P9 preferred to give up old habits "since I know that I have been playing badminton at a certain [high] level, it is frustrating to find myself being bad at it in the wheelchair. So maybe I would rather play tennis or basketball," whereas others saw their new life as a challenge to keep doing what they used to do. For example, regarding item 3, P1, who used to ski before his injury, felt that sit skiing would not be a challenge. When he tried it and realized how difficult it was for him, he was challenged to keep skiing:

> *I tried [sit ski], then it was so super hard.... So somewhere, it triggered me that I need to practice to get better at it.*

Regarding item 4, P9 reflected that his body did not react the same while running. As now he was not using his legs but his arms to move the wheelchair, it was harder for him to raise his pulse and feel the effects of exercise before his arms got tired. All of them had to learn to use a wheelchair, and the learning experience differed from person to person (item 5). For example, P4, one of the main organizers of a wheelchair community, mentioned that, when he started using a wheelchair, there was not much support but, in his community, they support others with basic wheelchair skills:

> *[W]e [practice] with balancing the wheelchair [on the back wheels] without dropping it back. [It] is very useful and hard skill.... It is very dangerous [as] you can hit your head.*

Everyday activities take more time based on the level of the injury and how it has influenced the body. For example, P7 mentioned that "I have to take care of myself and then work...it takes a lot of time," so she did not have much time to dedicate to exercising. P6, who worked at a rehabilitation center where people stayed for several weeks, said the following:

> *[R]ehabilitation is a lot about...you come here for 4, 6, 8, 10 weeks and you lay a foundation, but to get to be autonomous/independent. You need to continue when you get home because that's where you really need to put in the effort. And I tried to be clear with patients that this is the beginning. When you come home, that's where the struggle or the real-life starts.*

P6 underlined that, when one is in the rehabilitation clinic, they are free from daily responsibilities that may increase when they return to their daily life and take time from the exercise levels they need to sustain. Finally, P4 and P8 pointed out that, in addition to SCI, one develops more chronic conditions due to aging (eg, P8 kept track of his diet due to type 2 diabetes, and P4 tracked his blood pressure and related medication because he had a myocardial infarction).

To summarize, wheelchair users with SCI need to learn to know and adjust to their new bodies—bodies that are diverse by nature combined with the type of injury. When they learn basic wheelchair skills, they must deal with everyday life, which of course takes time for everyone, but for them, it can take time from their exercise. Exercise is a vital part of the high level of activity they need to sustain to be independent and healthy.

Reasons for Self-Tracking

Habitually, inpatients in rehabilitation clinics are told to document their activities or have them tracked during any type of rehabilitation or training related to their injury. However, few of our participants described this as tracking. For example, P9 mentioned the following:

> *If one is in rehab, one should keep track so they can see that the exercise is working...*

P7 remembered tracking during her injury-related training:

> *[Y]ou get a mentor that you can check up to after a couple of weeks, then you must make a new measure and see how you are doing. That was very useful.*

P8 reflected on the lack of technology and feedback during rehabilitation:

> *I think I get the idea [of tracking] from having to fill out all those forms when I was in the rehab center. Every week I have to write on paper, my weight, how much I lost when I went to the toilet, how much I ate on a particular day...and then the doctor concluded...[what I should change]...with the technology we have today just putting that on paper was not doing anything to me...I was just waiting for the outcome from the doctor.*

Our participants had different motivations for tracking. A total of 44% (4/9) of them (P1, P2, P5, and P9) mentioned that they had started tracking as they were practicing a sport. They tracked because their coach had told them to, or their coach did the

tracking for them. However, all of them kept tracking their behavior even after stopping exercising. P2 tracked variables related to swimming (eg, swimming style, time, and meters) as she was training for the Paralympics; nowadays, she tracks the activities she does through her Apple Watch as "it's so simple...the watch does it for me" and also uses it as a motivation to move—"it kind of reminds me to move" and "it keeps track, and it tells me you did a good [job] today." P1 mentioned his reaction the first time he was asked to track variables related to skiing (eg, weather conditions, exercise type, and timing) and how he came to keep tracking his activities:

> *Well, it was the trainer who said, you need to keep track on what you do. Why I've already done it. That was my answer. But then I started to do it. And then I realized that let's say after one season, I looked in the diary [and thought] oh, I need a lot of those things and [those things] didn't pay off. Maybe I should do less of those things since they don't pay off so much and maybe I should increase the things that pay off.*

P9 continued reflecting on his motivation for tracking his activities using his Apple Watch:

> *[I]t is not so much that I want to be in a national team or anything, it is more important that I can stay healthy and be a part of family activities and continue to do that. That is the most important thing to me.*

P3 mentioned keeping track of his training, especially when making changes to his training activities:

> *When you gonna train intensively for a month or something, I think it's interesting to keep track of your training. If you have a goal or race as well, which I don't right now.*

He reflected the following:

> *Yeah, I think most parts of tracking is for motivation for me at least, and then also, it's just interesting to see some data.*

P2 and P3 mentioned that they used tracking to increase motivation, as did P6:

> *[I]t's mostly motivation to see that to give myself a pat on the shoulder.*

Some participants (P1, P5, P6, and P9) tracked their activity before their injury as they were practicing different sports or participating in competitions. On the other hand, P4, P7, and P8 were not interested in tracking or were motivated to track only after their injury. P4 tracked his activity only as part of a game his colleague created—a point system that considered activities one did (eg, walking) and measurements of weight, waist, and chest. The person with the most points won. The point system was adjusted for P4 as he was in a wheelchair. P4 mentioned this game:

> *It was like, a challenge, you know...we challenge ourselves and even checked together "now I'm up to 2000 points" or whatever.*

P7, who had a higher level of injury, mentioned that she was never interested in tracking but that she kept track of her weight and the time it took her to move between places in her house:

> *[I]f you go to the bathroom, I got to make it in 10 minutes, [to] compete with myself to do with it as quickly as possible.... The only thing I track is my weight. My heart rate is no use because I can't get [much variation]*

P8 mentioned that he started feeling the need to track after his injury; before, he could simply see that he was muscular.

Summarizing, our participants performed a lot of tracking during their rehabilitation period to support the health care professionals following them, but the tracking was done on paper. Those training before their injury had a habit of tracking different variables and found value in tracking, but not everyone felt the need to track. In any case, the injury changed the way in which our participants tracked, either from not needing to track to needing to track or by changing the reasons for tracking and the variables they tracked.

Experiences With Self-Tracking Technologies and Tools

In this section, we will present the experiences of our participants with self-tracking technologies and tools through three subthemes: (1) use of self-tracking technologies, (2) tracking calories, and (3) trust in self-tracking technologies.

Self-Tracking Technology Use

The participants had experience using the Apple Watch, Runkeeper, Polar Beat, Pulse Polar Watch, and Nike+. While these technologies automate tracking and keep score of data, the participants also used manual means of tracking data about themselves. They mentioned using Microsoft Excel, a stopwatch, Lifesum, and paper as examples of manual means of self-tracking. One participant (P5) also mentioned that they "track in my mind." However, the most recurrent technologies in the responses were Apple Watch and Runkeeper.

Before their injury, only P6 and P9 used technology (Nike+ Running App) to track their activity. P1 kept a logbook with ski-related variables (he continued after the injury until he bought his Apple Watch), and P5's coach tracked his activity and gave him a schedule or advice accordingly. After their injury, participants P3, P5, P6, and P9 started using the Runkeeper app, which includes a wheelchair user mode. P3 and P5 used it to measure their speed and the distance they ran, and P6 and P9 used it to train for marathons or similar events. However, P9 mentioned that he had given up on Runkeeper because it did not work well with his Apple Watch:

> *Nike, that's actually what I used when [training for the] world run because it's, it's pre-installed in my Apple watch. So, when I got the Apple Watch, the first thing I did was download RunKeeper on my Apple watch. But I noticed that it wasn't attached to it with all its [features]. But there was also already a running up called Nike.*

P1, P2, and P9 were using Apple Watch to track their activities. Apple Watch includes an option for wheelchair users. In the

default mode, the Apple Watch tracks many activities (eg, biking, swimming, and skiing). However, if one sets it to the wheelchair user mode, only 4 activities become available: wheeling in- or outdoors and walking or running speed. P2 preferred to track her activities separately and used the default mode:

> *So most of the time I use the activities that are in the watch. I'm aware that they [calories] are not adapted for me and my muscle mass. That is not the important thing to me. For me, the important thing is that kind of keeps track and tells me that—okay, you did a good job.*

For P2, it was more important to track her activities and obtain positive feedback than to have access to accurate measurements of some of her activities.

The Apple Watch also displays 3 concentric rings: the inner blue ring shows how many times in the day one has moved (ie, changed position), the middle green circle shows how many minutes of brisk activity one has had, and the outer red circle indicates the active calories burned throughout the day. When P2 referred to the positive feedback she received, she meant how full these rings were. P1 and P9—who used the wheelchair mode on their watches—also looked at those rings for a quick daily overview. P9 mentioned the following:

> *[I] look at it but I don't pay too much attention to it.... I can see that I have done my full day's work and that's kind of the analysis I do, not so much else.*

Some of our participants used other apps and devices. For example, P8 had tried the Samsung Health app, but the measurements provided were not relevant to him. P1, one of the Apple Watch users, used a digital food diary (Lifesum) compatible with his watch to track his diet. P5, who was using Runkeeper, also used Polar Beat when training to measure his heart rate; however, he also considered the limitations of the technology. Specifically, he said the following:

> *I don't know if it's correct, but they mean that if you're a tetraplegic [injury level], your heart rate can't go over 120 Mm. And you can push as hard as you can. And you can't get over 120 But on the Polar-Beat that I [wear around the arm], I have been up to 136, I don't trust the one you put around [the chest] like this.*

This section shows that our participants were familiar with self-tracking tools. Many used or tried digital self-tracking technology after their injury, and some had done it even before. They tracked variables that could influence their sports performance or support them in understanding their body (eg, pulse and diet). They also tracked in less quantifiable ways, for example, what activities they did (eg, skiing or swimming) or the completion of goals set on their Apple Watch, which provided motivation. However, all participants mentioned calories and their measurement and consumption. Some expressed the need for a "lifestyle app" supporting them in healthy living as a whole—both in activity and diet tracking.

Tracking the Calories

Overview

When our participants talked about their experiences with self-tracking technology, the discussion always turned to calorie tracking or similar variables (such as weight and diet). Keeping a healthy lifestyle and not gaining weight is an issue for wheelchair users, and the impact of gaining weight on their life quality is tremendous. P9 was the most descriptive on how gaining weight could influence his life quality:

> *I just want one thing; I don't want to have to change my wheelchair into a wider one. If I get fatter, for example, it would be harder to make all these movements that I have to do in and out of the car every day, in and out of my bed every day in and out of the shower every day, or maybe every other day. Also, if because I'm getting fatter, if I have to get a wider wheelchair, I wouldn't get into certain doors, for example, in hotel rooms because my wheels are wider apart...I don't want that to happen and I'm ready to try any mechanical device to help me from not having to do that.*

This quote illustrates the potentially more serious consequences of gaining weight for wheelchair users and the need to track the calorie burn and, if possible, intake. P5 mentioned that "in our world, calories are very important," and P1 commented on weight gain and aging:

> *[W]eight is a problem all the time, especially now when I am getting a bit older, and it is easier to gain weight than losing it.*

Even P7, who seemed uninterested in tracking her activities and did not use any self-tracking technology, mentioned that "tracking it's not so important for me" but "I track weight because it is important for us." When our participants discussed tracking their calories, they mostly talked about calorie burn (P1, P2, P3, P5, P6, P8, and P9). Only 44% (4/9) of the participants (P1, P3, P4, and P5) mentioned calorie intake or diet, and only 11% (1/9; P7) mentioned weight (which can be perceived as a balance between calorie intake and burn).

Calorie Burn

As wheelchair users with SCI have less muscle mass than a walking person, they consume calories differently. P1 mentioned the following:

> *It is really hard for a paraplegic [injury level] to burn calories. It takes so much more time to burn the same amount as you [a full-muscle ability person]. If you're out jogging for one hour, I must wheel maybe for three.*

Therefore, to burn enough calories, they need to spend more time exercising.

Few commercial self-tracking technologies have their features adjusted to wheelchair users. Runkeeper and Apple Watch were 2 technologies mentioned by our participants that consider wheelchair users in some of their features. Our participants did not trust the self-tracking technology they used regarding calories and, instead, used other variables as a proxy for calorie

burn. Consequently, they managed their calories by staying active or using self-tracking technology as a motivation to stay active. P6 used self-tracking technology (FitNotes and Runkeeper) to keep a high level of activity in addition to motivating himself:

> *[One session] wasn't good enough, next week, I'm doing two sessions. And I'm taking the time [for an extra session] by putting something else out of my schedule. Or if, last week, I trained five times. I tell myself I did something good. Continue!*

P5, who used the Runkeeper app, underlined that he checked the distance and speed provided by the app but not the calories despite mentioning that calories were very important to him. He also discarded the calories that his wearable showed (Polar Beat, which measures heartbeat and calories), saying that "I don't care about the calories because I know...the calories are all wrong." P3 also mentioned heartbeat as an indicator for understanding that his training had an impact:

> *[H]eartbeat is always good because if your heartbeat races, you're doing something.*

He also reflected on a noncommercial smartwatch he had tried as part of another study that measured steps, finding steps not to be a good indicator or proxy for understanding his activity levels:

> *I do not know if I am moving a lot, I have more steps on my watch, but I do not know how it works and I do not know what it reduces it and what it does not.*

Similarly, P8 reflected that the visualization of steps in commercial self-tracking technology as an activity proxy was irrelevant to him:

> *Samsung health (smartphone) counts my steps which is not connected with the reality...Like FitBit.*

Apple Watch is one of the self-tracking technologies that has adjusted some of its features for wheelchair users (eg, wheeling in- or outdoors and walking or running pace); however, only P1 trusted and followed their calorie burn. In contrast, P9 pointed out the following:

> *I have my Apple Watch on when I do [strength exercises]. I guess it's not smart...it tracks some of the movements I make as a push in my wheelchair. It was the same with that would that pole pooling...it doesn't track it as something else than kind of push with the wheelchair.*

Thus, the watch counted his calories based on the strokes he did to move his wheelchair, and he argued that, regardless of what exercise he did, the watch would always measure only wheeling in- or outdoors and walking or running pace but not the actual exercise. Therefore, the calories could not be estimated based on a different activity but only on the strokes he did, so he did not pay attention to them but used the watch mainly for motivation. P2, another Apple Watch user, said the following:

> *[A]ll other activities are in there, but they are not adapted for people with no full muscle ability. With that said, I had kind of decided that it doesn't really matter if I spent 300 calories or 500 calories.*

She added that she used the watch to motivate herself to move. Thus, she preferred to use the Apple Watch in the default mode rather than in the wheelchair mode, favoring the ability to track different activities over a more accurate calorie estimation.

Some participants also used nontechnological means to measure their calories manually. For example, P1, who was part of a medical research project on SCI, became aware of the connection between his energy consumption and his pulse and took action based on that:

> *[B]efore being part of the project about energy consumption for a disabled person, I had no way to determine how much energy I consume during the day but after that, I made charts because I have a really strong correlation between pulse and how much energy I consume.*

To summarize, a person with SCI consumes calories at a slower pace than a full-muscle, able-bodied person and needs to be more active to burn the same amount of calories. Most commercial technology currently in the market does not account for this difference, which leads to apps incorporating the wheelchair symbol without adjusting the calorie burn for this type of user or apps calculating the calorie burn correctly but only for limited activities. Thus, our participants estimated their calorie burn based on how many exercise sessions they did or their heart rate increase. Self-tracking technology was used mainly for motivation, and steps were not perceived as a good proxy for activity tracking or calorie burn.

Calorie Intake

Weight is impacted not only by calorie expenditure but also by calorie intake. For example, P4 said that "Exercises are good but if you don't stop eating too much, you will never lose weight." Just like P3 adjusted his exercise knowledge to the new situation, P5 needed to adjust what he knew about diet. Before the injury, P5 had a personal trainer who planned his weekly exercise and diet, for example, to gain muscle. However, P5 mentioned that he found that he could no longer follow what he had learned from his trainer after the injury:

> *When I ended up in a wheelchair, I was used to having the [diet and exercise] paper in the back of my head. But every other person in a wheelchair said you can't eat as much as you did before. So today I'm eating small portions but more often. I don't do breakfast, lunch, dinner.*

P5 had to adapt his weight management after becoming a wheelchair user by avoiding the traditional food distribution throughout the day.

Even though participants were interested in counting their calorie intake, only P1 used a dietary app, called Lifesum. He mentioned that the intake of calories was also important and that it would be interesting to have a daily calorie goal and take note of the type of diet one follows, if applicable. P1 said the following:

> *An app that could measure what I eat, and I can put in parameters like, [amount of] calories [that] are ok for the day, and any kind of diet I follow.*

Another participant, P7, felt that she had control over her diet owing to a course organized for people with SCI that she took after her injury:

> *[In this course] you start tracking everything, they check your weight and height, how you train and exercise, how you feel—everything. And then you get a plan and learn a lot [about] bad calories, and you know exactly how much you are going to eat and how to change...you get a mentor that you can check in with after a couple of weeks, then you [must] make a new measure or everything and see how you are doing. That was very useful.*

Here, instead of using a tracking tool, the user learned how to manage their calorie intake in an educational setting, when episodic measurements were made, and then made sense of them with the support of their mentor.

The participants expressed that calorie intake was as important as calorie burn. They described that they had to adjust their eating habits after the injury to fit their new bodies. Even though only P1 used self-tracking technology to track their diet, other participants mentioned that they had to undergo training or go through an adjustment period to learn how much they could eat by tracking their diet.

Trust in Self-Tracking Technology

P1, P2, P3, P5, P6, and P9 used self-tracking technology to track their activities but did not trust this technology, especially regarding the calories. P1 trusted his Apple Watch because "Apple bought a survey conducted in the US that was about paraplegics, working [out], and how much calories [they burn] with which workload. And they put that algorithm in the watch. For example, let's say a paraplegic person has a 30% slower or lower metabolism compared to able-bodied, that is in consideration in this app." He compared his watch to Runkeeper:

> *I'm out wheeling with 100 pulses; I get depressed because it's like 210 calories. But if you go to RunKeeper, and you have been jogging for one hour, you have 700 calories. So, if you use that app, you will be so fooled and tricked by it. Because it is hard for a paraplegic to burn calories.*

P1 made a comment on commercial apps using the wheelchair symbol deceptively:

> *[Y]ou have apps on the market RunKeeper, or whatever, you download it because it has a wheelchair symbol [thinking] Oh, yeah, this is a cool company, they even bother to [consider] wheelchair [users]. Yeah, but people don't understand. That's a chart with calories from an able-bodied person. They just replace the running person symbol with the wheelchair symbol.*

P2 was dissatisfied that, as a wheelchair user, she could only track 4 activities on the Apple Watch. Therefore, she preferred to use the watch without the wheelchair user setting to keep track of the different activities rather than accurately tracking the calories—as the quote in the previous theme shows. Despite this, P2 stated that they were an "Apple family" and that she trusted Apple because "it's people I know that are interested in technology that buy all the gear and try all the gear out that's Apple Watch is good I'm buying kind of trusted them Yeah. And my husband recommended and said that it would be worth it [now] that I had already started to add more exercise into my life, and I thought that could be a good fun thing to try." In this case, what made the technology more trustworthy was not the actual technological features but the social context. On the other hand, P9 did not trust his Apple Watch to accurately calculate the calories because it tracked everything as strokes on a wheelchair regardless of the type of activity—as his quote in the previous theme showed. However, he mentioned that he trusted his Apple Watch more than an app on his phone as he could not see how an app could track accurately:

> *[T]here was a wheelchair feature that I can put [in the RunKeeper], instead of running or cycling, I put it on the chair. But for example, it didn't track my pushes, because it can't track my pushes, it seems it's only on the phone. So, I think it helps that it's on my Apple watch on my hand to be able to keep track of [the pushes]. So, if I only have the phone in my, pouch that I have on my sitting pillow, or I have it on my arm higher, it doesn't track my movement in the same way.*

P9 expressed his distrust of Runkeeper because it did not provide the user with an explanation of how the app measures the calories of a wheelchair user.

Similarly, P3 and P5 mentioned that they did not trust the technology they used (Runkeeper and Polar Beat) regarding the calories it presented them with. Specifically, P5 mentioned the following:

> *Now it's like, if I go for a walk, it shows how far and the pace and even it [RunKeeper] shows the calories. But I do not think the calories are exact, so I don't care so much about the calories.*

He explained that, apart from calories, "I trust just the RunKeeper because it's like a map, a GPS. You have walked three kilometers. You got the pace for each kilometer." This quote illustrates that, despite P5 distrusting some features of the app (calories), he trusted others that were more easily understandable. He added the following:

> *So for now, I'm pretty happy with the Polar-Beat it is just the calories are not...and it's not reliable.*

Although the user considered himself satisfied with the Polar Beat use, he chose not to rely on the calorie measurement.

Summarizing, trust in self-tracking technology (or the lack of it) is influenced by the perception of the participants regarding how the technology works and how the development company accounts for different bodies related to calorie burn.

Discussion

Principal Findings

This section recaps our contribution followed by a short discussion on trust and transparency related to self-tracking technology. Then follows a section with the implications of commercial self-tracking technology, which concludes with directions for future research.

To address the question of how wheelchair users with SCI experience self-tracking and commercial self-tracking technology in their everyday lives, we interviewed 9 wheelchair users with SCI who had experience with self-tracking or self-tracking technology. The last theme, *experiences with self-tracking technology and tools* in Figure 1 divided into 3 subthemes, shows that current self-tracking technology fails to accommodate wheelchair users' needs concerning calorie management in 2 ways. First, the current commercial technology calorie calculation is based on a body with full muscle ability and, therefore, is inaccurate for people with SCI (subtheme: *calorie tracking*), and second, the users do not trust the calories in the self-tracking technology (subtheme: *trust in self-tracking technology*). Nevertheless, the wheelchair users we studied continue to use self-tracking technology to motivate them to stay active and to monitor variables such as heartbeat, speed, and activity schedule as proxies for their energy consumption (subtheme: *self-tracking technology use*).

This research has limitations due to the small number of interviews and the positioning of the research in a technologically advanced country (Sweden), which also has 1 of the 7 lowest gaps in unemployment between people with and without disabilities in the EU. If the study had been conducted in a different context with, for example, lower technological literacy and use or lower employability of people with disabilities, the results may have differed, particularly on the number of people who have used self-tracking technology. Finally, all of our participants (9/9, 100%) were White. If our participants had had a darker skin tone, the results may have differed, particularly in relation to the *trust in self-tracking technology* theme as self-tracking technology is even less accurate on darker skin [53]. Nevertheless, this study adds to the body of knowledge by validating previous studies on the different variables that wheelchair users want to see. It expands the current body of knowledge by focusing on long-lived experiences of self-tracking and contemporary commercial self-tracking technology, especially on calorie management—the most important factor influencing the life quality of wheelchair users with SCI.

Our research contributes to the scientific community as follows. It extends the work by Malu and Findlater [5] on fitness trackers on wheelchairs by interviewing wheelchair users with SCI who had lived, daily experience with tracking their activity or self-tracking technologies. In line with the work by Malu and Findlater [5], our findings in the *calorie tracking* subtheme indicate that one of the most important values to be measured is calories (intake and burn), followed by distances, GPS data, and pulse, as indicated by the *self-tracking technology use* subtheme. However, our findings in the *calorie burn* subtheme indicate that steps are not a good variable to measure for wheelchair users, which contradicts the study by Malu and Findlater [5] but is in line with the work by Carrington et al [6] and Helle and Rosenbeck Gøegb [39].

Our research, specifically the *self-tracking technology use* subtheme, is in line with the work by Helle Rosenbeck Gøegb[39] regarding the tracking of distance and time. In addition, the *trust in self-tracking technologies* subtheme is partially in line with the results of Helle and Rosenbeck Gøegb [39] showing a lack of trust in technology related to some variables (calorie burn and training intensity). Our participants, similarly to those in the study by Helle and Rosenbeck Gøegb [39], trusted the self-tracking technology regarding their speed measurements and were interested in these measurements; however, neither the participants in the aforementioned study nor our participants trusted calorie burn tracking (see the *trust in self-tracking technology* subtheme). Some of our participants showed a lack of trust in training intensity as well (similarly to those in the study by Helle and Rosenbeck Gøegb [39]); for example, smartwatches counted every activity as a stroke on a chair without distinguishing it from upper-body strength training (see the *trust in self-tracking technology* subtheme). Finally, our findings in the *reasons for self-tracking* theme suggest that being able to track different activities and calorie burn and keep track of exercise schedules, as well as seeing activity goals being reached, was motivational for our participants to keep the active lifestyle they needed, which is in line with the findings of past research [27,28,54].

Trust and Transparency

Participants' trust in self-tracking technologies was heavily influenced by their understanding and perception of how these devices worked. For example, Apple Watch users often trusted the device more than phone apps, believing that the watch's ability to track arm movements made it inherently more accurate than a phone carried in a pocket. However, this trust was conditional and limited; participants expressed confidence in the calorie estimates only for the 4 wheelchair-specific activities provided by the device. Regarding other activities in which calorie expenditure was approximated using these predefined categories, they expressed skepticism about accuracy. This led to varying strategies—some participants attempted to fit all their activities into 1 of the 4 wheelchair-specific categories to achieve more reliable calorie estimates, whereas others discarded calorie tracking entirely and used the watch in nonwheelchair mode to access a broader range of features.

Previous work on trust in self-tracking technology has revealed that users often have a *mental model* of how apps or devices work and calculate data [55]. Trust in self-tracking technologies, particularly in the context of wheelchair users with SCI, hinges significantly on the transparency of how data are collected, processed, and presented. Participants in our study exhibited a nuanced understanding of their devices, often forming mental models to rationalize the data's validity. However, the perceived "black box" nature of calorie calculations—whether on the Apple Watch or Runkeeper—fostered skepticism. This reinforces previous findings that transparency regarding algorithms and data processing is critical for trust [56]. On the

basis of this, we suggest that designers prioritize user education by including detailed explanations of their algorithms, especially when adapting generic metrics such as calorie burn for niche user groups, such as wheelchair users. This aligns with broader calls [57,58] for designing technologies that foster informed trust.

Furthermore, trust in technology is not isolated from social factors [59]. Apple Watch users' trust was partially rooted in social recommendations and previous brand experiences. This suggests that trust may extend beyond technical accuracy to include the brand's perceived reliability and commitment to inclusion. Investigating the interplay between social networks, brand trust, and technology adoption among wheelchair users could provide further insights into how collective validation influences individual trust decisions.

Distrust in calorie tracking was not merely technical but also emotional, linked to fears of inaccuracy impacting critical life decisions (eg, managing weight to maintain mobility). This illustrates that trust is not just a cognitive process but a deeply affective one. Incorporating participatory design with wheelchair users at every stage of the design and development process—testing, iteration, and after launch—can mitigate such emotional concerns. By co-designing solutions, developers can address both the technical and psychological dimensions of trust.

We observed how distrust often led participants to bypass calorie metrics altogether, choosing instead to focus on other variables or manual methods. This self-empowerment reflects resilience but also signals a failure of the technology to deliver on its promise. Therefore, tools must enhance user agency by aligning design goals with real-world constraints and expectations.

Finally, we observed a mismatch between representation and actual design. Runkeeper users appreciated the app's reliance on GPS and speed data, which they found straightforward and comprehensible. However, despite this, they distrusted the app's calorie calculations even when using its wheelchair mode and described feeling misled by its implied inclusivity. The use of wheelchair symbols created an initial perception of inclusivity but, ultimately, resulted in feelings of deception among participants due to inaccurate calorie calculations. This reflects a broader issue in which symbols and icons signaling accessibility and inclusiveness are in fact merely superficial. Misrepresentations such as these can diminish trust but also risk alienating the very communities that the technologies claim to support. Implications such as these are discussed further in the next section.

Implications for Commercial Self-Tracking Technology for Wheelchair Users With SCI

The most important need and valuable variable for our participants (subtheme: calorie tracking), and of the participants of previous research with wheelchair users, was the accuracy of calorie consumption [6,39,60]. Previous research [6,39,60] and our participants indicated that commercial fitness and wellness technology that uses the wheelchair symbol does not necessarily calculate accurately the calories burned by wheelchair users (subtheme: *trust in self-tracking technologies*). Research [6,39,60] also indicates that specific technology for wheelchair users needs to be developed, personalized, and tailored to their needs, which underlines the importance of tailoring mainstream self-tracking technology to the user group that it intends to include [61]. Furthermore, there is extensive research on wheelchair user–specific technology for self-tracking health and well-being [27,28,54], as well as exploring current technology with wheelchair users to inspire designs of wheelchair user–specific fitness and wellness technology [5]. Even though a big part of the aforementioned literature focuses on assistive technology or technology specifically for wheelchair users, mainstream self-tracking technology that aims to be inclusive toward wheelchair users can take advantage of what assistive technology research can offer (eg, databases and artificial intelligence models) [30,31], as well as what the medical field offers, which already has some physical activity guidelines for wheelchair users [14,32,34,62].

On the other hand, the production of new wheelchair-specific self-tracking technology (which is not assistive technology) divides the market into wheelchair users and non–wheelchair users. Recent research [41] indicates that, by excluding specific populations from the development of commercial products, the incorporation of accessibility requirements into general consumer products comes to a halt. In addition, special technology could cost more as it targets a specific population. Regarding wheelchair users, special technology may not be a financially viable solution as they have a higher risk of financial issues [42]. Working with affordable materials to create new technology for a specific population can also lead to them not using it because it could be stigmatizing [63]. From 2025 onward, digital services and products in the EU market should be accessible according to the EAA [10,49]. Our findings show that the self-tracking technology currently in the market is not prepared to be accessible to wheelchair users with SCI. This technology is not trusted when it comes to calorie calculation. Our participants described how they used proxy variables to estimate their calorie consumption and activity level. They expressed the need for a self-tracking technology that accurately calculates their calorie intake and burn during a variety of activities. Many of the issues with self-tracking technology that uses the wheelchair symbol can be solved by involving wheelchair users in the design and development phase. This involvement will make the technology more relevant to the people it aims to include (ie, wheelchair users) and open the market to broader audiences [64-66].

On the basis of the literature presented in this paper and our findings, we see a need for future research and development regarding inclusive self-tracking technology for wheelchair users with SCI supported and inspired by assistive technology literature.

First, there is a need to explore ways to make affordable technology already used by wheelchair users closer to their needs and make it more accurate regarding calorie burn (ie, track a variety of activities and calorie intake). Our study can be seen as a starting point for this exploration of commercial fitness self-tracking technology and its use by wheelchair users with SCI, complementing previous literature on similar subjects [5-7,11].

Second, we argue that there is a need to confront and acknowledge the limitations of the methods used to conduct similar research and alternate them for the research and practice to be more inclusive. According to Kabir et al [40], data collection and design methods may be inaccessible to some populations, leading researchers to exclude them as they perceive them as "difficult to reach and research." Instead, they suggest having a palette of methods and adjusting them based on the participants' abilities [40]. Similarly, Moon et al [37] provide a comprehensive toolbox for more inclusive user research when designing with wheelchair users in mind. In addition, Li et al [30] and Huang et al [31] provide freely available concrete tools (ie, databases and code) for developing an inclusive self-tracking technology to be used by anyone wanting to make self-tracking technology more inclusive.

Third, regardless of all the support that this paper and the literature provide, the inclusion of wheelchair users as an integral part of the design process of the self-tracking technology targeting this demographic should be seen as mandatory for the product to be relevant to the new audience (ie, wheelchair users) [65,66].

Future Research

Our next steps are a series of participatory workshops in which we—together with people with SCI and the support of the SCI center we partnered with—identify further tracking needs and co-design a first version of an app that could support them in learning about their new body in different stages of their postinjury phase. In the workshops, we plan to include the physiotherapist who cooperates with us to include the health care perspective without compromising the power balance. His participation will also create an environment in which the participants can learn more about SCI and how the health care system perceives SCI and them as patients. To create a link for the communication between health care professionals and people with SCI, we will use the connections of our partner at Karolinska Institute to investigate the needs of health care professionals (ie, which of the data tracked by people with SCI are relevant to the health care professionals and in what form to fit their practice when consulting with their patients).

Acknowledgments

The authors thank all the people who dedicated their time to participate in the interviews. Without their contribution and their rich experiences, this study would have been impossible. The authors thank their collaborators from the spinal cord injury center that supported this research and helped them find participants for this study.

Authors' Contributions

VM led the research and the writing process of the paper. VM designed and conducted the interviews and the initial thematic analysis with support from KC. AW and MR participated in the later stages of the thematic analysis and contributed to the overall writing of the paper. TH contributed to the writing of the paper and was responsible for the medical accuracy of the paper regarding spinal cord injury.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The final themes resulting from the thematic analysis in a tabular form.
[PDF File (Adobe PDF File), 28 KB - humanfactors_v12i1e65207_app1.pdf]

References

1. International perspectives on spinal cord injury. World Health Organization. URL: https://www.who.int/publications/i/item/international-perspectives-on-spinal-cord-injury [accessed 2024-04-29]
2. Sprigle S, Sonenblum S, O'Brien JJ. Transitioning a research tool into a consumer product: the wheelchair in-seat activity tracker. In: Proceedings of the 2019 Conference on Design of Medical Devices. 2019 Presented at: DMD '19; April 15-18, 2019; Minneapolis, MN p. 1-5. [doi: 10.1115/DMD2019-3211]
3. Büyüktür AG, Ackerman MS, Newman MW, Hung PY. Design considerations for semi-automated tracking self-care plans in spinal cord injury. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17; May 23-26, 2017; Barcelona, Spain p. 183-192 URL: https://dl.acm.org/doi/10.1145/3154862.3154870 [doi: 10.1145/3154862.3154870]
4. Ahad N, Sonenblum SE, Davenport MA, Sprigle S. Validating a wheelchair in-seat activity tracker. Assist Technol 2022 Sep 03;34(5):588-598 [FREE Full text] [doi: 10.1080/10400435.2021.1893237] [Medline: 33617402]
5. Malu M, Findlater L. Toward accessible health and fitness tracking for people with mobility impairments. In: Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2016 Presented at: PervasiveHealth '16; May 16-19, 2016; Cancun, Mexico p. 170-177 URL: https://dl.acm.org/doi/10.5555/3021319.3021344 [doi: 10.4108/eai.16-5-2016.2263329]

6. Carrington P, Chang K, Mentis H, Hurst A. "But, I don't take steps": examining the inaccessibility of fitness trackers for wheelchair athletes. In: Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility. 2015 Presented at: ASSETS '15; October 26-28, 2015; Lisbon, Portugal p. 193-201 URL: https://dl.acm.org/doi/10.1145/2700648.2809845 [doi: 10.1145/2700648.2809845]
7. Ungurean OC, Vatavu RD. “I gave up wearing rings:” insights on the perceptions and preferences of wheelchair users for interactions with wearables. IEEE Pervasive Comput 2022 Jul 1;21(3):92-101. [doi: 10.1109/MPRV.2022.3155952]
8. Sloane M. Inequality is the name of the game: thoughts on the emerging field of technology, ethics and social justice. In: Proceedings of the Weizenbaum Conference 2019 on Challenges of Digital Inequality - Digital Education, Digital Work, Digital Life. 2019 Presented at: EFTESJ '19; May 16-17, 2019; Berlin, Germany p. 1-9 URL: https://tinyurl.com/2b2r7sjj [doi: 10.34669/wi.cp/2.9]
9. Shew A. Ableism, technoableism, and future AI. IEEE Technol Soc Mag 2020 Mar;39(1):40-85. [doi: 10.1109/mts.2020.2967492]
10. European accessibility act. European Parliament. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32019L0882 [accessed 2023-03-10]
11. Bernard RM, Seijas V, Davis M, Volkova A, Diviani N, Lüscher J, et al. Mobile health self-management support for spinal cord injury: systematic literature review. JMIR Mhealth Uhealth 2023 Apr 26;11:e42679 [FREE Full text] [doi: 10.2196/42679] [Medline: 37099372]
12. Spinal locomotor circuits: organization and repair after injury. European Commission. 2011. URL: https://cordis.europa.eu/project/id/201144/es [accessed 2024-04-29]
13. Holmlund T, Ekblom-Bak E, Franzén E, Hultling C, Wikmar LN, Wahman K. Energy expenditure in people with motor-complete paraplegia. Spinal Cord 2017 Aug 4;55(8):774-781. [doi: 10.1038/sc.2017.27] [Medline: 28374809]
14. Holmlund T, Ekblom-Bak E, Franzén E, Hultling C, Wahman K. Energy expenditure after spinal cord injury in people with motor-complete tetraplegia or motor-complete paraplegia. Spinal Cord 2018 Mar 14;56(3):274-283. [doi: 10.1038/s41393-017-0024-4] [Medline: 29238094]
15. Krassioukov A. Autonomic function following cervical spinal cord injury. Respir Physiol Neurobiol 2009 Nov 30;169(2):157-164. [doi: 10.1016/j.resp.2009.08.003] [Medline: 19682607]
16. Piaggi P, Thearle MS, Bogardus C, Krakoff J. Lower energy expenditure predicts long-term increases in weight and fat mass. J Clin Endocrinol Metab 2013 Apr;98(4):E703-E707 [FREE Full text] [doi: 10.1210/jc.2012-3529] [Medline: 23418317]
17. Piaggi P, Vinales KL, Basolo A, Santini F, Krakoff J. Energy expenditure in the etiology of human obesity: spendthrift and thrifty metabolic phenotypes and energy-sensing mechanisms. J Endocrinol Invest 2018 Jan 24;41(1):83-89 [FREE Full text] [doi: 10.1007/s40618-017-0732-9] [Medline: 28741280]
18. Piaggi P. Metabolic determinants of weight gain in humans. Obesity (Silver Spring) 2019 May 23;27(5):691-699 [FREE Full text] [doi: 10.1002/oby.22456] [Medline: 31012296]
19. Chen Y, Henson S, Jackson AB, Richards JS. Obesity intervention in persons with spinal cord injury. Spinal Cord 2006 Feb 16;44(2):82-91. [doi: 10.1038/sj.sc.3101818] [Medline: 16103891]
20. Chen Y, Cao Y, Allen V, Richards JS. Weight matters: physical and psychosocial well being of persons with spinal cord injury in relation to body mass index. Arch Phys Med Rehabil 2011 Mar;92(3):391-398. [doi: 10.1016/j.apmr.2010.06.030] [Medline: 21238950]
21. Lupton D. The diverse domains of quantified selves: self-tracking modes and dataveillance. Econ Soc 2016 Apr 15;45(1):101-122. [doi: 10.1080/03085147.2016.1143726]
22. Global report on assistive technology. World Health Organization. URL: http://apps.who.int/bookorders [accessed 2025-01-27]
23. Redelings MD, Lee NE, Sorvillo F. Pressure ulcers: more lethal than we thought? Adv Skin Wound Care 2005 Sep;18(7):367-372. [doi: 10.1097/00129334-200509000-00010] [Medline: 16160463]
24. Motahar T, Ghosh I, Wiese J. Identifying factors that inhibit self-care behavior among individuals with severe spinal cord injury. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA p. 1-16 URL: https://dl.acm.org/doi/10.1145/3491102.3517658 [doi: 10.1145/3491102.3517658]
25. Amann J, Fiordelli M, Brach M, Bertschy S, Scheel-Sailer A, Rubinelli S. Co-designing a self-management app prototype to support people with spinal cord injury in the prevention of pressure injuries: mixed methods study. JMIR Mhealth Uhealth 2020 Jul 09;8(7):e18018 [FREE Full text] [doi: 10.2196/18018] [Medline: 32673241]
26. Amann J, Fiordelli M, Scheel-Sailer A, Brach M, Rubinelli S. Opportunities and challenges of a self-management app to support people with spinal cord injury in the prevention of pressure injuries: qualitative study. JMIR Mhealth Uhealth 2020 Dec 09;8(12):e22452 [FREE Full text] [doi: 10.2196/22452] [Medline: 33295876]
27. Hiremath SV, Amiri AM, Thapa-Chhetry B, Snethen G, Schmidt-Read M, Ramos-Lamboy M, et al. Mobile health-based physical activity intervention for individuals with spinal cord injury in the community: a pilot study. PLoS One 2019 Oct 15;14(10):e0223762 [FREE Full text] [doi: 10.1371/journal.pone.0223762] [Medline: 31613909]

28. Lakshminarayanan R, Canori A, Ponnada A, Nunn M, Read MS, Hiremath SV, et al. Exploring opportunities to improve physical activity in individuals with spinal cord injury using context-aware messaging. Proc ACM Hum Comput Interact 2022 Nov 11;6(CSCW2):1-27. [doi: 10.1145/3555628]
29. Mo W, Singh A, Holloway C. Exploring information needs for tracking to support using wheelchairs in everyday life. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. 2024 Presented at: CHI EA '24; May 11-16, 2024; Honolulu, HI p. 1-7 URL: https://dl.acm.org/doi/10.1145/3613905.3650933 [doi: 10.1145/3613905.3650933]
30. Li Y, Mollyn V, Yuan K, Carrington P. WheelPoser: sparse-IMU based body pose estimation for wheelchair users. In: Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility. 2024 Presented at: ASSETS '24; October 27-30, 2024; St. John's, NL p. 1-17 URL: https://dl.acm.org/doi/10.1145/3663548.3675638 [doi: 10.1145/3663548.3675638]
31. Huang W, Ghahremani S, Pei S, Zhang Y. WheelPose: data synthesis techniques to improve pose estimation performance on wheelchair users. In: Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems. 2024 Presented at: CHI '24; May 11-16, 2024; Honolulu, HI p. 1-25 URL: https://dl.acm.org/doi/10.1145/3613904.3642555 [doi: 10.1145/3613904.3642555]
32. Hoevenaars D, Holla JF, Postma K, van der Woude LH, Janssen TW, de Groot S. Associations between meeting exercise guidelines, physical fitness, and health in people with spinal cord injury. Disabil Rehabil 2023 Mar 20;45(6):1030-1037 [FREE Full text] [doi: 10.1080/09638288.2022.2048910] [Medline: 35306932]
33. Tweedy SM, Beckman EM, Geraghty TJ, Theisen D, Perret C, Harvey LA, et al. Exercise and sports science Australia (ESSA) position statement on exercise and spinal cord injury. J Sci Med Sport 2017 Feb;20(2):108-115. [doi: 10.1016/j.jsams.2016.02.001] [Medline: 27185457]
34. Martin Ginis KA, van der Scheer JW, Latimer-Cheung AE, Barrow A, Bourne C, Carruthers P, et al. Evidence-based scientific exercise guidelines for adults with spinal cord injury: an update and a new guideline. Spinal Cord 2018 Apr;56(4):308-321. [doi: 10.1038/s41393-017-0017-3] [Medline: 29070812]
35. Li Y, Li F, Carrington P. Breaking the “inescapable” cycle of pain: supporting wheelchair users’ upper extremity health awareness and management with tracking technologies. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany p. 1-17 URL: https://dl.acm.org/doi/10.1145/3544548.3580660 [doi: 10.1145/3544548.3580660]
36. Motahar T, Wiese J. A review of personal informatics research for people with motor disabilities. Proc ACM Interact Mob Wearable Ubiquitous Technol 2022 Jul 07;6(2):1-31. [doi: 10.1145/3534614]
37. Moon NW, Baker PM, Goughnour K. Designing wearable technologies for users with disabilities: accessibility, usability, and connectivity factors. J Rehabil Assist Technol Eng 2019 Aug 13;6:2055668319862137 [FREE Full text] [doi: 10.1177/2055668319862137] [Medline: 35186318]
38. Glasheen E, Domingo A, Kressler J. Accuracy of Apple Watch fitness tracker for wheelchair use varies according to movement frequency and task. Ann Phys Rehabil Med 2021 Jan;64(1):101382 [FREE Full text] [doi: 10.1016/j.rehab.2020.03.007] [Medline: 32335302]
39. Helle T, Rosenbeck Gøegb K. Wheelchair users’ experiences with and need of activity trackers. In: Proceedings of the 2019 International Conference on Advancement of Assistive Technology in Europe. 2019 Jul 17 Presented at: ESERA '19; August 26-30, 2019; Bologna, Italy p. 1 URL: https://www.ucviden.dk/en/publications/wheelchair-users-experiences-with-and-need-of-activity-trackers-3 [doi: 10.3233/tad-190001]
40. Kabir KS, Alsaleem A, Wiese J. The impact of spinal cord injury on participation in human-centered research. In: Proceedings of the 2021 ACM Designing Interactive Systems Conference. 2021 Presented at: DIS '21; June 28-July 2, 2021; Virtual Event p. 1902-1914. [doi: 10.1145/3461778.3462122]
41. Eskytė I. Disabled people’s vulnerability in the European single market: the case of consumer information. J Consum Policy 2019 Jul 19;42(4):521-543. [doi: 10.1007/S10603-019-09422-3]
42. Disability statistics - financial situation - statistics explained. Eurostat. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Disability_statistics_-_financial_situation#Struggling_to_make_ends_meet [accessed 2023-01-20]
43. Kouprie M, Visser FS. A framework for empathy in design: stepping into and out of the user's life. J Eng Des 2009 Oct;20(5):437-448. [doi: 10.1080/09544820902875033]
44. Bennett CL, Rosner DK. The promise of empathy: design, disability, and knowing the "other". In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 1-13 URL: https://dl.acm.org/doi/10.1145/3290605.3300528 [doi: 10.1145/3290605.3300528]
45. Lag (2003:460) om etikprövning av forskning som avser människor. Sveriges Riksdag. URL: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som_sfs-2003-460/ [accessed 2024-04-29]
46. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
47. Patton MQ, Welch JK. Qualitative evaluation and research methods. Mod Lang J 1992;76(4):543. [doi: 10.2307/330063]
48. Williams M, Moser T. The art of coding and thematic exploration in qualitative research. Int J Manag Rev 2019;15(1):45-55 [FREE Full text]

49. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
50. Raguindin PF, Muka T, Glisic M. Sex and gender gap in spinal cord injury research: focus on cardiometabolic diseases. A mini review. Maturitas 2021 May;147:14-18 [FREE Full text] [doi: 10.1016/j.maturitas.2021.03.004] [Medline: 33832642]
51. E-tjänster. Svenskarna och internet. URL: https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2024/e-tjanster/#betaltjanster-och-swish [accessed 2025-02-07]
52. Majority of persons with disabilities locked out of quality employment. European Disability Forum. URL: https://www.edf-feph.org/majority-of-persons-with-disabilities-locked-out-of-quality-employment/ [accessed 2025-02-07]
53. Colvonen PJ, DeYoung PN, Bosompra NO, Owens RL. Limiting racial disparities and bias for wearable devices in health science research. Sleep 2020 Oct 13;43(10):zsaa159 [FREE Full text] [doi: 10.1093/sleep/zsaa159] [Medline: 32893865]
54. Mason L, Dickinson P, Gerling K, De Angeli A. Design goals for playful technology to support physical activity among wheelchair users. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19; May 4-9, 2019; Glasgow, UK p. 1-12 URL: https://dl.acm.org/doi/10.1145/3290605.3300262 [doi: 10.1145/3290605.3300262]
55. Zhang Y, Trace CB. The quality of health and wellness self‐tracking data: a consumer perspective. J Assoc Inf Sci Technol 2021 Oct 14;73(6):879-891. [doi: 10.1002/ASI.24591]
56. Kizilcec RF. How much information?: effects of transparency on trust in an algorithmic interface. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI '16; May 7-12, 2016; San Jose, CA p. 2390-2395 URL: https://dl.acm.org/doi/10.1145/2858036.2858402 [doi: 10.1145/2858036.2858402]
57. Felzmann H, Fosch-Villaronga E, Lutz C, Tamò-Larrieux A. Towards transparency by design for artificial intelligence. Sci Eng Ethics 2020 Nov 16;26(6):3333-3361. [doi: 10.1007/s11948-020-00276-4]
58. Vorm ES, Combs DJ. Integrating transparency, trust, and acceptance: the intelligent systems technology acceptance model (ISTAM). Int J Hum Comput Interact 2022 May 20;38(18-20):1828-1845. [doi: 10.1080/10447318.2022.2070107]
59. Riegelsberger J, Sasse MA, McCarthy JD. The mechanics of trust: a framework for research and design. Int J Hum Comput Interact 2005 Mar;62(3):381-422. [doi: 10.1016/J.IJHCS.2005.01.001]
60. Malu M, Chundury P, Findlater L. Exploring accessible Smartwatch interactions for people with upper body motor impairments. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-12 URL: https://dl.acm.org/doi/10.1145/3173574.3174062 [doi: 10.1145/3173574.3174062]
61. Warms C. Physical activity measurement in persons with chronic and disabling conditions: methods, strategies, and issues. Fam Community Health 2006;29(1 Suppl):78S-88S [FREE Full text] [doi: 10.1097/00003727-200601001-00012] [Medline: 16344640]
62. Holmlund T, Ekblom-Bak E, Franzén E, Hultling C, Wahman K. Intensity of physical activity as a percentage of peak oxygen uptake, heart rate and Borg RPE in motor-complete para- and tetraplegia. PLoS One 2019 Dec 3;14(12):e0222542 [FREE Full text] [doi: 10.1371/journal.pone.0222542] [Medline: 31794548]
63. Carrington P, Hurst A, Kane SK. Wearables and chairables: inclusive design of mobile input and output techniques for power wheelchair users. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, ON p. 3103-3112 URL: https://dl.acm.org/doi/10.1145/2556288.2557237 [doi: 10.1145/2556288.2557237]
64. Waller S, Bradley M, Hosking I, Clarkson PJ. Making the case for inclusive design. Appl Ergon 2015 Jan;46 Pt B:297-303. [doi: 10.1016/j.apergo.2013.03.012] [Medline: 23538129]
65. Bødker S, Kyng M. Participatory design that matters—facing the big issues. ACM Trans Comput Hum Interact 2018 Feb 13;25(1):1-31. [doi: 10.1145/3152421]
66. Wilkinson CR, De Angeli A. Applying user centred and participatory design approaches to commercial product development. Des Stud 2014 Nov;35(6):614-631. [doi: 10.1016/j.destud.2014.06.001]

Abbreviations

EAA: European Accessibility Act
EU: European Union
SCI: spinal cord injury

Edited by M Waite; submitted 12.08.24; peer-reviewed by A Panahi, A Castan; comments to author 22.10.24; revised version received 07.02.25; accepted 07.03.25; published 15.04.25.

Please cite as:
Mylonopoulou V, Cerna K, Weilenmann A, Rost M, Holmlund T
Experiences of Wheelchair Users With Spinal Cord Injury With Self-Tracking and Commercial Self-Tracking Technology ("In Our World, Calories Are Very Important"): Qualitative Interview Study
JMIR Hum Factors 2025;12:e65207
URL: https://humanfactors.jmir.org/2025/1/e65207
doi:10.2196/65207
PMID:

The Chinese Version of the DigiHealthCom (Digital Health Competence) Instrument for Assessing Digital Health Competence of Health Care Professionals: Translation, Adaptation, and Validation Study

Lu Gao[1], BS; Meilian Chen[1], BS; Jingxin Wei[1], BS; Jinni Wang[1], MSN; Xiaoyan Liao[2,3], Prof Dr

[1]School of Nursing, Southern Medical University, Guangzhou, China

[2]Nanfang Hospital, Southern Medical University, Guangzhou, China

[3]School of Nursing, Guangzhou Medical University, Panyu Campus, No. 1 Xinzao Road, Panyu District, Guangzhou, China

Corresponding Author:
Xiaoyan Liao, Prof Dr
Nanfang Hospital, Southern Medical University, Guangzhou, China

Abstract

Background: Digital health competence is increasingly recognized as a core competence for health care professionals. A comprehensive evaluation of knowledge, skills, performance, values, and attitudes necessary to adapt to evolving digital health technologies is essential. DigiHealthCom (Digital Health Competence) is a well-established instrument designed to assess digital health competence across diverse health care professionals.

Objective: This study aimed to translate and culturally adapt DigiHealthCom into simplified Chinese (Mandarin) and verify its reliability and validity in assessing digital health competence of Chinese health care professionals.

Methods: DigiHealthCom was translated into Chinese following the guideline proposed by its original developers. The cultural adaptation involved expert review and cognitive interviewing. Internal consistency, test-retest reliability, content validity, convergent validity, discriminant validity, and factor structure were examined. Item analysis tested item discrimination, item correlation, and item homogeneity. Internal consistency was assessed using Cronbach α, and test-retest reliability was measured using the intraclass correlation coefficient. Content validity was assessed through both item and scale content validity indices. Convergent validity was measured by the Average Variance Extracted and Composite Reliability, while discriminant validity was measured by the heterotrait-monotrait ratio. A five-dimension model of DigiHealthCom was confirmed using confirmatory factor analysis.

Results: The finalized Chinese version of the DigiHealthCom was completed after addressing differences between the back-translations and the original version. No discrepancies affecting item clarity were reported during cognitive interviewing. The validation process involved 398 eligible health care professionals from 36 cities across 15 provinces in China, with 43 participants undergoing a retest after a 2-week interval. Critical ratio values (range 16.05 - 23.77, P<.001), item-total correlation coefficients (range 0.69 - 0.89), and Cronbach α if the item deleted (range 0.91 - 0.96) indicated satisfactory item discrimination, item correlation, and item homogeneity. Cronbach α for dimensions and the scale ranged from 0.94 to 0.98, indicating good internal consistency. The intraclass correlation coefficient was 0.90 (95% CI 0.81 - 0.95), indicating good test-retest reliability. Item content validity index ranged from 0.82 to 1.00, and the scale content validity index was 0.97, indicating satisfactory content validity. Convergent validity (average variance extracted: 0.60 - 0.79; composite reliability: 0.94 - 0.95) and divergent validity (heterotrait-monotrait ratio: 0.72 - 0.89) were satisfactory. Confirmatory factor analysis confirmed a well-fit five-dimension model (robust chi-square to df ratio=3.10, comparative fit index=0.91, Tucker-Lewis index=0.90, incremental fit index=0.91, root-mean-square error of approximation=0.07, standardized root-mean-square residual=0.05), with each item having a factor loading exceeding 0.40.

Conclusions: The Chinese version of DigiHealthCom has been proved to be reliable and valid. It is now available for assessing digital health competence among Chinese health care professionals. This assessment can be used to guide health care policy makers and educators in designing comprehensive and implementable educational programs and interventions.

(JMIR Hum Factors 2025;12:e65373) doi:10.2196/65373

KEYWORDS

competence; digital health; health care professionals; instrument; reliability; validity

Introduction

The World Health Organization (WHO) defines digital health as the field that involves the development and use of digital technologies to improve health outcomes [1]. In 2019, the WHO released the world's first guidelines for digital health interventions, outlining 10 ways to use digital health technologies to enhance health and primary services [2]. This concept extends beyond eHealth to include a wide array of smart devices and connected equipment, encompassing digital technologies such as the Internet of Things, big data, artificial intelligence, and robotics [2]. The release of these guidelines was followed by the Global Strategy for Digital Health (2020 - 2025), which emphasizes prioritized digital health strategies for global health care development [1]. China has incorporated measures for the development of a digitally healthy nation in the 14th Five-Year Plan [3]. With the global market projected to grow from US $211 billion in 2022 to US $809.2 billion by 2030 [4], digital health care is recognized as a rapidly expanding sector. Digital health has emerged as a significant trend in the evolution of global health care services. The digital transformation that the health care sector is currently undergoing is redefining the roles and responsibilities of health care professionals [5,6], creating an urgent need for digital health competence among them.

Given increasing prominence of digital health in global health care landscape, it is critical for health care professionals to possess sufficient digital health competence. Digital health competence is increasingly recognized as one of the core competencies for health care professionals, which would enable them to design and evaluate the impact of digital solutions on patient care and determine the best way to implement digital solutions in their work [7,8]. Although patients are becoming more accepting of and motivated to use digital health care services and tools, health care professionals face a digital skills shortage that impedes the adoption of digital solutions [9-11]. Inadequate digital health competence may lead to negative experiences and frustration with technology adoption among these professionals [12]. There is a strong association between health care professionals' digital health competence and their willingness to use such tools [13,14]. Studies have highlighted that the acceptance of digital health technologies by health care professionals significantly influences the adoption of digital solutions and emphasizes the critical role of digital health competence in ensuring patient safety [12,14]. Therefore, assessing digital health competence is essential to effectively providing digital health care solutions to the public.

Previous studies have attempted to explore digital health competence among health care professionals but have struggled to define it comprehensively due to the evolving nature of digital technologies [15]. Existing assessment tools primarily focus on informatics competence, digital health literacy, or skills related to the application of digital technologies. Examples include the Digital Health Literacy Instrument (DHLI, 2017) [16], the eHealth literacy questionnaire (eHLQ, 2018) [17], and the Nursing Digital Application Skill Scale (NDASS, 2024) [3]. The eHLQ is designed for eHealth user, especially individuals with low digital health literacy and those with chronic conditions [17]. The NDASS targets nurses' digital application skills in clinical settings [3]. Digital health literacy reflects users' knowledge and skills within their cultural, social, and institutional context [17]. Competence entails an integrative understanding of the knowledge, skills, performance, values, and attitudes essential for the effective execution of a given task [18]. Therefore, it is crucial to comprehensively evaluate the knowledge, skills, performance, and attitudes required for various health care professionals to adapt to the evolving digital health technologies.

Developed and validated among Finnish health care professionals in 2022, the DigiHealthCom (Digital Health Competence) instrument offers a more comprehensive scope than existing tools and is applicable to a wide range of health care professionals. In addition to assessing competence in using digital solutions and information and communication technology (ICT), it also explores previously unaddressed domains that reflect future requirements, such as acceptance of digital solutions, human-centered remote counseling, and ethical competence concerning digital solutions [19]. Furthermore, the instrument has been used to explore digital health competence profiles and associated factors in 817 health care professionals from 9 organizations in Finland [20]. It has been translated into 15 languages, and a large-scale international cross-sectional study on the digital health competence of health care professionals is currently in progress. Our team is part of this collaborative research effort. This study aimed to culturally adapt and validate the Chinese version of DigiHealthCom for Chinese health care professionals.

Methods

Study Design

A cross-sectional study was conducted.

Participants

Participants were recruited between May 2023 and April 2024 via convenience sampling. Recruitment posters with QR codes were disseminated on social networks specific to health care professionals. In addition, health care professionals attending local academic conferences were invited to participate. The inclusion criteria for participants were (1) employment within a health care organization and (2) consent to participate. The exclusion criteria were (1) individuals who were retired or had less than 1 year of work experience; (2) health care students; (3) individuals who completed the questionnaire in less than 150 seconds, as this indicated random clicking; or (4) individuals who displayed erratic response patterns.

According to the thumb rule, the sample size should be 5 to 10 times the number of items. With an anticipated 10% rate of invalid responses, a minimum of 231 samples was necessary. Finally, a total of 398 cases were included in the study. To assess the test-retest reliability of the instrument, 43 participants who provided contact information completed the questionnaire again after a 2-week interval.

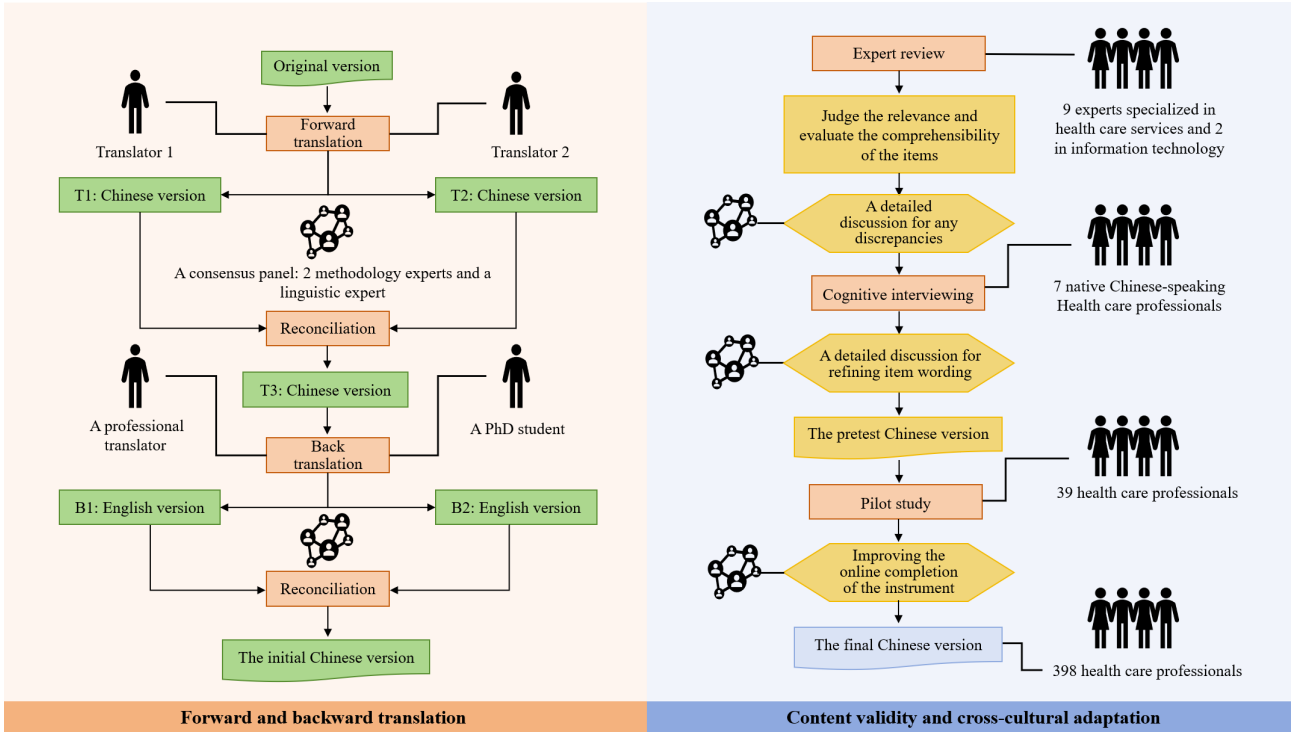
Instrument

The DigiHealthCom instrument comprises 5 domains, comprising a total of 42 items—competence in human-centered remote counseling (16 items), digital solutions as part of work (9 items), competence in ICT (5 items), competence in using and evaluating digital solutions (8 items), and ethical competence related to digital solutions (4 items). Each item was rated on a 4-point Likert scale (1=completely disagree, 2=partially disagree, 3=partially agree, and 4=completely agree). For each domain, a mean value of ≤2.49 indicated low competence, 2.50 - 3.49 indicated intermediate competence, and ≥3.50 indicated high competence [20]. DigiHealthCom has been validated among health care professionals across tertiary, primary, and private health care settings (n=817), demonstrating satisfactory internal consistency (Cronbach α=0.80 - 0.97) and content validity (item content validity index [I-CVI]: 0.77 - 1.00; S-CVI/Ave: 0.94) [19].

Translation and Cross-Cultural Adaptation

To ensure a high-quality Chinese translation of DigiHealthCom, a rigorous translation and cross-cultural adaptation process was followed [21]. Figure 1 illustrates the translation and cross-cultural adaptation process.

Figure 1. The translation and cross-cultural adaptation process of developing the Chinese version of the DigiHealthCom (Digital Health Competence) instrument.



Forward and Backward Translation

In total, 2 bilingual nursing professionals, who are native Chinese speakers (a nursing expert and a nursing graduate student), independently translated the English version of DigiHealthCom into Chinese, resulting in 2 forward translations (T1 and T2). A consensus panel reviewed these translations for conceptual equivalence, clarity, and comprehensibility, producing a reconciled version (T3). Subsequently, the T3 version was back-translated into English by a professional translator and a PhD student unfamiliar with the original version, both knowledgeable about Chinese and English-speaking cultures. The consensus panel reviewed and resolved discrepancies between the back-translations and the original version, finalizing the initial Chinese version. The consensus panel comprised 2 nursing researchers experienced in scale development and a linguistic expert.

Expert Review

To assess the content validity of the Chinese version of DigiHealthCom, 11 experts were invited to evaluate the relevance of its dimensions and items using a 4-point ordinal scale (1 [not relevant], 2 [weakly relevant], 3 [strongly relevant], and 4 [very relevant]). The expert panel consisted of 9 health care specialists and 2 IT experts. In addition, expert evaluated the comprehensibility of the items. The consensus panel, initially involved in the translation process, conducted a detailed discussion to resolve any potential discrepancies.

Cognitive Interviewing

Cognitive interviewing was used to evaluate the clarity and cultural suitability of the initial Chinese version. A total of 7 native Chinese speakers, including 2 doctors, 4 nurses, and 1 IT technician, were recruited. Participants were briefed on the study's objectives and methods before interviews, and their consent was obtained.

The first author, trained in qualitative research methods, conducted the interviews in a meeting room. Participants completed the Chinese version of DigiHealthCom independently and engaged in cognitive interviews. Interviewers evaluated whether participants found the items relevant to their condition and if they encountered any understanding difficulties. Field notes were taken. Modifications were deemed necessary if at

least 1 participant (1) found an item difficult to understand, (2) demonstrated mostly or completely inaccurate comprehension of an item, or (3) provided feedback indicating the need for improvements, especially regarding cultural relevance. The consensus panel determined whether to retain or alter item wording following expert review and cognitive interviewing. Discrepancies affecting item clarity were resolved to create the pretest version of the Chinese DigiHealthCom version. Specific item modifications were detailed in Table S1 in Multimedia Appendix 1.

Pilot Study

A pilot study was conducted with 39 health care professionals from a general hospital in Guangzhou, China. The pilot study confirmed that no further modifications were required. The final Chinese version of DigiHealthCom consists of 5 dimensions and 42 items.

Data Collection

Data were collected using the Wen Juan Xing (a Chinese web questionnaire platform) [22]. Respondents accessed the questionnaire by scanning a QR code or clicking a link, with 1 response allowed per IP address to prevent duplicates. The questionnaire platform performed completeness checks before submission. The Chinese version of DigiHealthCom was presented alongside a sociodemographic questionnaire. The questionnaire consisted of 2 pages and 73 items. Sex, age, education, service area, professional license, work experience, type of organization, and frequency of patient work were collected. Participants who voluntarily chose to provide contact information participated in a retest after a 2-week interval. An introduction outlined the study's purpose and provided questionnaire instructions.

Statistical Analysis

Item analysis was used to evaluate item discrimination, item correlation, and item homogeneity. For item discrimination, respondents were classified into high-score (top 27%) and low-score (bottom 27%) groups based on their total scores. An independent *t* test determined whether each item could significantly distinguish between these groups. Items with a critical ratio ($|t|$) less than 3.0 were considered for exclusion [23]. In addition, item homogeneity and item correlation were tested using Cronbach α if the item was deleted and corrected item-total correlation coefficient. Items with a corrected item-total correlation coefficient below 0.40 were considered for exclusion [23]. Cronbach α and Intraclass correlation coefficient (ICC) were used to assess the internal consistency and test-retest reliability of the instrument. A Cronbach α≥0.70 indicated good internal consistency, while ICC>0.70 indicated good time stability [23]. The I-CVI and the Scale CVI/Average (S-CVI/Ave) were used to evaluate content validity of the instrument. I-CVI≥0.78 and S-CVI/Ave≥0.90 indicate satisfactory content validity [24].

Construct validity was evaluated using confirmatory factor analysis (CFA) with a robust chi-square to df ratio (χ^2/df) less than 3, Tucker-Lewis Index (TLI), Incremental Fit Index (IFI), and Comparative Fit Index (CFI) greater than 0.90, and root-mean-square error of approximation (RMSEA) and standardized root-mean-square residual (SRMR) less than 0.08, indicating an acceptable data-model fit [25]. The average variance extracted (AVE) and composite reliability (CR) were used to assess convergent validity, with AVE greater than 0.50 and CR greater than 0.70 indicating good convergent validity [26]. Discriminant validity was performed using the heterotrait-monotrait ratio (HTMT), with a correlation matrix value <0.90 considered good [26].

Furthermore, participants were divided into 2 groups based on geographic location for the sensitivity analysis. The DigiHealthCom scores were compared to evaluate potential selection bias. The absence of a significant difference in digital health competence between the groups indicates that selection bias is unlikely in the study sample. Statistical analysis was conducted using IBM SPSS (version 27.0), IBM AMOS (version 29.0), and Smart PLS (Version 4.1.0.0; GmbH Corp). A *P* value of less than .05 was considered statistically significant.

Ethical Considerations

Ethical approval was obtained from the Ethical Committee at Nanfang Hospital, Southern Medical University, China (NFEC-2023 - 165). This research adhered to the principals of the Declaration of Helsinki. All participants provided informed consent and voluntarily completed the web questionnaire. Participants had the option to skip questions, review, and delete their responses. Participants had the right to withdraw from the survey at any time. Those who completed it received a random monetary reward ranging from CNY 2 to 5 (US $0.28 to 0.69). Participants' rights and researcher's contact information were provided on the first page of the web survey. Minimal sociodemographic was collected to maintain ethical standards. Participant information was kept confidential and anonymous. The CHERRIES (Checklist for Reporting Results of Internet E-Surveys; Checklist 1) was used to enhance the transparency of the study [27].

Results

Characteristics of the Participants

Initially, 431 participants were recruited for this study. Furthermore, 33 participants were subsequently excluded due to not meeting the inclusion criteria (n=7), having less than 1 year work experience (n=4), completing the questionnaire in under 150 seconds (n=2), and exhibiting erratic response patterns (n=20). Finally, 398 eligible health care professionals were included. Figure 2 illustrates the participant enrollment flowchart. The participants enrolled from 36 cities across 15 provinces, as identified by IP addresses. Among the participants, 249 (62.6%) were recruited from Guangzhou, a megacity in Guangdong province that contains 6125 registered medical facilities. As shown in Table 1, 364 (91.5%) participants were female, 386 (97%) worked in health care services, and 357 (89.7%) were nurses. The participants had an average of 13.7 (SD 9.4) years of work experience, with 281 (70.6%) working directly with patients for at least 5 days per week. Participant characteristics are summarized in Table 1.

Figure 2. The flowchart for the enrollment process of health care professionals.

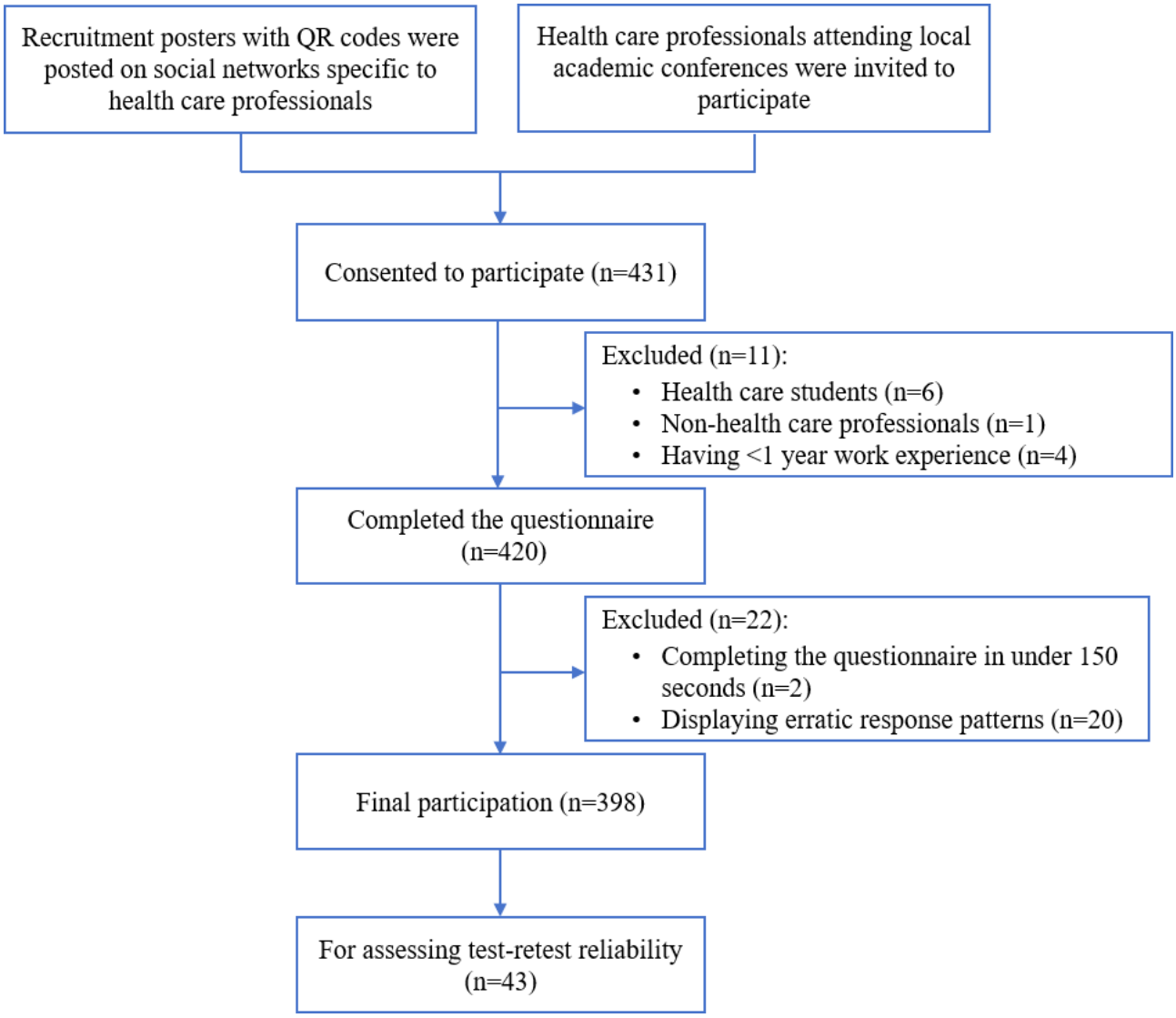


Table . Demographic characteristics of the participants (N=398).

Variables		Value
Sex, n (%)		
	Female	364 (91.5)
	Male	34 (8.5)
Age (years), mean (SD)		36.1 (8.8)
Education, n (%)		
	Junior vocational qualification	58 (14.6)
	Bachelor's degree	273 (68.6)
	Master's degree	53 (13.3)
	Doctoral degree	14 (3.5)
Service area, n (%)		
	Health care service	386 (97)
	Social service	4 (1)
	Rehabilitation service	7 (1.8)
	Others	1 (0.3)
Location, n (%)		
	Southern China	351 (88.2)
	Northern and Western China	47 (11.8)
Type of organization, n (%)		
	Tertiary hospital	248 (62.3)
	Secondary hospital	36 (9)
	Community health care center	76 (19.1)
Professional license, n (%)		
	Nurse	357 (89.7)
	Doctor	24 (6)
	Midwife	10 (2.5)
	Others[a]	7 (1.8)
Working experience (years), mean (SD)		13.7 (9.4)
Patient work, n (%)		
	Daily (at least 5 days a week)	281 (70.6)
	Weekly (1-4 days per week)	62 (15.6)
	Monthly (a few times a month)	17 (4.3)
	Rarely (a few times in several months)	23 (5.8)
	I do not currently work with patient	15 (3.8)
Full-time, n (%)		398 (100)

[a]Including physiotherapist, paramedical technician, and pharmacist.

Results of Item Analysis

Item analysis showed a significant difference between high-score and low-score groups. The critical ratio values for all items were above 3.0 (range 16.05 - 23.77, P<.001; Table 2), indicating excellent item discrimination. All corrected item-total correlation coefficients exceeded 0.4 (Table 2). The Cronbach α if the item deleted (range 0.91 - 0.96) were acceptable (Table 2).

Table . Item analysis and content validity of the 42 items in the DigiHealthCom (Digital Health Competence; N=398).

Item	Critical ratio	CITC[a]	CID[b]	I-CVI[c]
RC1[d]	16.05	0.69	0.96	1
RC2	17.27	0.70	0.96	1
RC3	17.29	0.71	0.96	1
RC4	19.43	0.77	0.96	1
RC5	17.10	0.77	0.96	1
RC6	20.34	0.79	0.96	1
RC7	19.90	0.80	0.96	1
RC8	22.99	0.79	0.96	0.82
RC9	19.33	0.76	0.96	1
RC10	21.91	0.81	0.96	1
RC11	19.26	0.76	0.96	1
RC12	20.84	0.80	0.96	0.91
RC13	19.89	0.80	0.96	1
RC14	20.80	0.74	0.96	1
RC15	19.02	0.73	0.96	0.91
RC16	18.71	0.75	0.96	0.91
DS1[e]	21.53	0.77	0.96	1
DS2	20.63	0.81	0.95	1
DS3	19.82	0.84	0.95	1
DS4	21.53	0.86	0.95	1
DS5	22.19	0.89	0.95	0.91
DS6	22.00	0.86	0.95	1
DS7	22.57	0.87	0.95	0.91
DS8	20.62	0.71	0.96	0.91
DS9	23.77	0.81	0.95	0.91
ICT1[f]	20.62	0.84	0.92	1
ICT2	18.64	0.88	0.91	1
ICT3	22.03	0.88	0.91	1
ICT4	18.67	0.81	0.93	1
ICT5	18.88	0.75	0.94	0.82
UE1[g]	20.48	0.82	0.95	1
UE2	21.89	0.82	0.95	1
UE3	23.04	0.84	0.95	1
UE4	22.25	0.87	0.95	0.91
UE5	21.33	0.83	0.95	1
UE6	18.19	0.78	0.95	0.91
UE7	23.55	0.87	0.95	0.91
UE8	20.52	0.82	0.95	1
EC1[h]	21.40	0.84	0.92	1
EC2	20.79	0.81	0.93	0.91
EC3	20.48	0.89	0.91	1

Item	Critical ratio	CITC[a]	CID[b]	I-CVI[c]
EC4	21.44	0.86	0.91	1

[a]CITC: corrected item-total correlation.

[b]CID: Cronbach α if item deleted.

[c]I-CVI: item content validity index.

[d]RC: human-centered remote counseling competence.

[e]DS: digital solutions as part of work.

[f]ICT: information and communication technology competence.

[g]UE: competence in using and evaluating digital solutions.

[h]EC: ethical competence related to digital solution.

Internal Consistency and Test-Retest Reliability

As shown in Table 3, internal consistency (Cronbach α=0.94 - 0.98) and test-retest reliability were good (ICC=0.90; 95% CI 0.81 - 0.95).

Table . Internal consistency and test-retest reliability of the Chinese version of the DigiHealthCom (Digital Health Competence).

Dimension	Cronbach α (N=398)	ICC[a] (N=43)	95% CI of ICC
RC[b]	0.96	0.80	0.63-0.89
DS[c]	0.96	0.90	0.81-0.95
ICT[d]	0.94	0.79	0.62-0.89
UE[e]	0.96	0.80	0.63-0.89
EC[f]	0.94	0.81	0.65-0.90
Total	0.98	0.90	0.81-0.95

[a]ICC: intraclass correlation coefficient.

[b]RC: human-centered remote counseling competence.

[c]DS: digital solutions as part of work.

[d]ICT: information and communication technology competence.

[e]UE: competence in using and evaluating digital solutions.

[f]EC: ethical competence related to digital solutions.

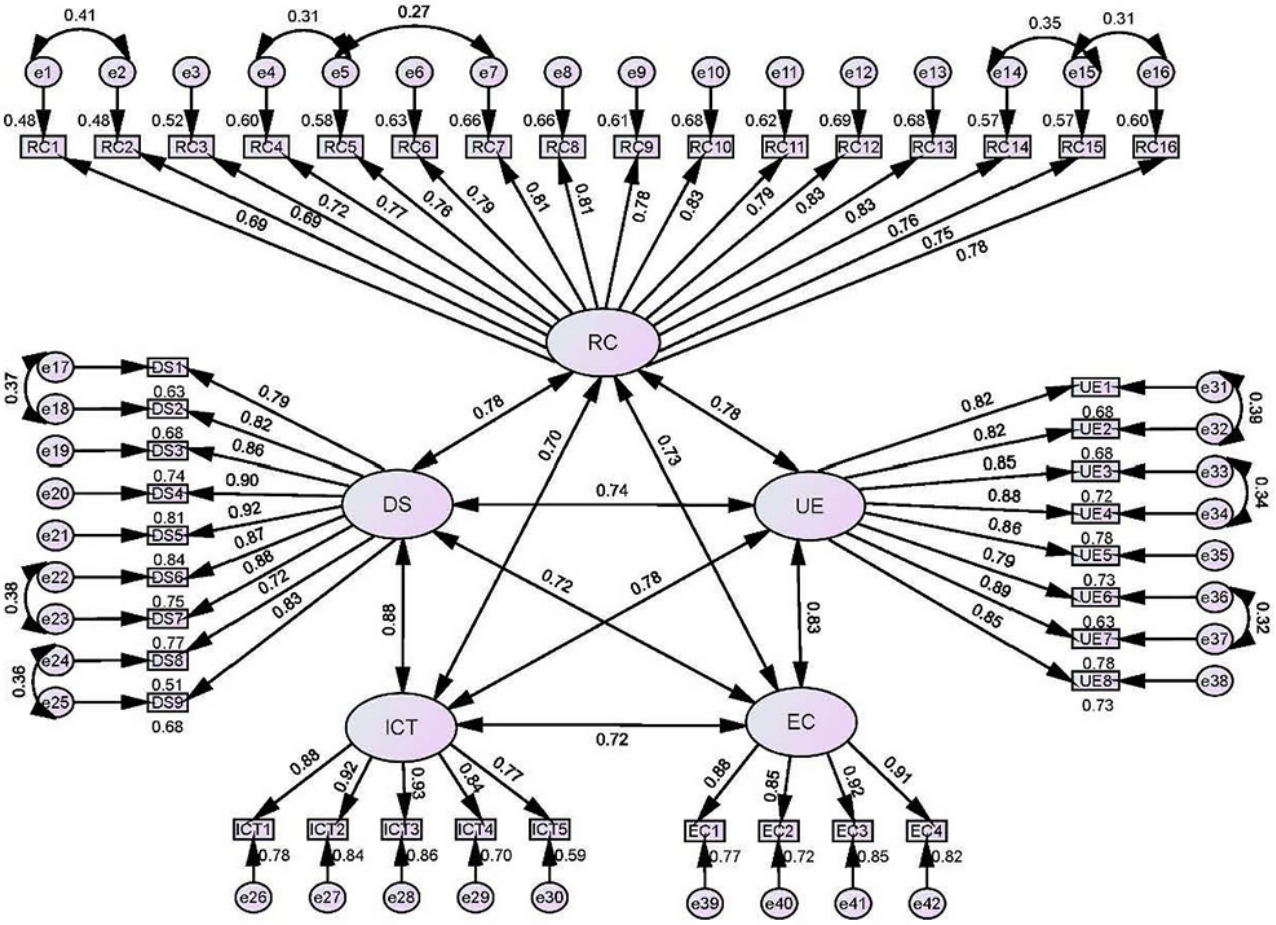
Content Validity

To test content validity, 11 experts evaluated the relevance of the items. The S-CVI/Ave for the instrument was 0.97, and the I-CVI ranged from 0.82 to 1.00 (Table 2), indicating satisfactory content validity.

Construct Validity

As illustrated in Figure 3, the results of CFA supported the 5-factors structure (digital solutions as part of work; ethical competence related to digital solutions; ICT competence; human-centered remote counselling competence; competence in using and evaluating digital solutions). The indices, including χ^2/df (3.10), CFI (0.91), TLI (0.90), IFI (0.91), RMSEA (0.07), and SRMR (0.05), indicated an acceptable model fit. The factor loadings of items ranged from 0.69 to 0.93 (Figure 3). The AVE values for each dimension exceeded 0.50 (range: 0.60 - 0.79), with the CR values >0.70 (range: 0.94 - 0.95), indicating excellent convergent validity. The HTMT values within the matrix were less than 0.90 (range: 0.72 - 0.89), indicating good divergent validity. These results collectively demonstrated satisfactory construct validity of the instrument.

Figure 3. The confirmatory factor model of the Chinese version of the DigiHealthCom (Digital Health Competence; N=398). DS: digital solutions as part of work; EC: ethical competence related to digital solutions; ICT: information and communication technology competence; RC: human-centered remote counselling competence; UE: competence in using and evaluating digital solutions.



Sensitivity Analysis

Participants were divided into 2 groups based on geographic location: Southern China and Northern and Western China. The DigiHealthCom scores were compared to evaluate potential selection bias. As illustrated in Table S3 in Multimedia Appendix 2, no significant difference was found in DigiHealthCom scores between the 2 groups (P>.05), indicating that selection bias is unlikely in the study sample.

Discussion

Principal Findings

In this study, we translated and culturally adapted the DigiHealthCom instrument into Chinese and assessed its reliability and validity among Chinese health care professionals. The Chinese version of DigiHealthCom demonstrated satisfactory internal consistency, test-retest reliability, content validity, and construct validity. To our knowledge, this study represents the first effort to validate a comprehensive Chinese tool for measuring digital health competence among health care professionals.

In this study, each item of the Chinese version of DigiHealthCom was translated and back-translated strictly following the dual direct-to-back translation model [21] to ensure alignment in semantic, conceptual, and content with the original English version. And then, we conducted expert review involving 11 experts in nursing and IT to examine content validity of the Chinese version. High content validity indices imply that the instrument provides a broad enough range of content to allow conclusions about the targeted construct. We also conducted cognitive interviewing and a pilot study to ensure its clarity and comprehensibility [28]. Issues related to semantic validation and understanding identified in the initial Chinese version were addressed after receiving feedback from the expert review and cognitive interviewing, leading to improvements such as clarifying complex terms and vague definitions. Item analysis revealed good differentiation and high correlations between the items in this study. Furthermore, this instrument underwent rigorous testing for internal consistency, test-retest reliability, and construct validity, exhibiting excellent internal consistency, time stability, and construct validity, resonating with the original study [19]. The types of reliability and validity assessed in this study represent essential psychometric properties for a measurement instrument.

Discrepancies between existing instruments largely stem from differing conceptual frameworks. The DHLI measures 7 individual skills: operation, navigation, information searching, evaluating reliability, determining relevance, adding self-generated content, and protecting privacy [16]. The eHLQ comprises 7 dimensions, addressing users' attributes, their interaction with technologies, and their experience with systems

[17]. Our CFA findings confirm a 5-factor structure, consistent with the exploratory factor analysis findings of the initial study [19]. Beyond competence in using and evaluating digital solutions and ICTs, the DigiHealthCom instrument addresses additional essential domains that were previously unaddressed, that is acceptance of digital solutions, human-centered remote counseling, digital interaction skills with patients and interprofessional teams, and ethical competence related to digital solutions [19]. Notably, acceptance of digital solutions significantly influences health care professionals' adoption of digital solutions [13,14]. Competence in person-centered remote consultations is crucial for fostering patient engagement and improving accessibility to equitable digital health services [8]. Furthermore, proficient digital interaction skills facilitate effective coordination in digital settings, thereby supporting decision-making and optimizing treatment plans [8]. As digital solutions become more prevalent, attention to data privacy and information security has intensified [29,30]. Ethical competence related to digital solutions ensures appropriate management of patient information, fostering trust in digital health technologies and enhancing the safety of digital health services [20,31]. These domains provide a unified theoretical framework for comprehensively measuring digital health competencies among various professionals [8,9,19], which are essential for delivering high-quality digital health services to meet future demands.

The Finnish health care system is renowned for its universality, robust public funding, and centralized digital infrastructure [32]. Regional variations exist in Finland; for example, northern districts have fewer professionals specializing in remote counseling and the integration of digital solutions compared with southern areas [20,33]. Strategies have been implemented to enhance health care professionals' digital health competence and improve education related to health care digitalization, including technology to support client engagement, digital services integrated into nursing work, and considerations of safety and ethics in the digital environment [20,34]. In contrast, China experiences regional disparities in digital health infrastructure and access to digital health services, especially in rural areas, despite the rapid growth of digital health services [35]. In addition, system compatibility across hospitals remains an unsolved issue [35]. These factors may influence the development of digital health competence among health care professionals. A cross-sectional study conducted in a central Chinese province reported that 49.9% (1690/3386) of clinical nurses demonstrate low telehealth readiness [36]. Consequently, evaluating, addressing, and reducing regional disparities in health care professionals' digital health competence is vital for promoting equity in health care service provision. Validating a comprehensive digital health competence measurement instrument for health care professionals is imperative to address challenges posed by diverse health care environments, such as China's. The Chinese version of DigiHealthCom is now ready for application in a wide range of settings and various health care professionals. Researchers, educators, health care providers, and policy makers can use it to evaluate digital health competence among diverse health care professionals, and develop digital health training curricula and policies based on the assessment.

Limitations

This study acknowledges several limitations. First, the primary participants were nurses. Future research should involve a broader spectrum of health care professionals. Second, although recruited from diverse health care institutions, the participants mainly came from southern China, which may introduce selection bias due to regional disparities in digital health infrastructure. To address this, participants were divided into 2 groups (Southern China versus Northern and Western China) for sensitivity analysis, and their DigiHealthCom scores were compared. We did not find significant differences in digital health competence between the 2 groups, indicating that selection bias is unlikely in the study sample. Future studies could benefit from employing a stratified sampling method. Third, the absence of a standardized method for assessing digital health competence among health care professionals hindered the evaluation of criterion validity. Assessments of digital health competence predominantly rely on subjective self-reports, which might be susceptible to response bias. However, relying exclusively on performance-based assessments or other objective measures to evaluate competence also presents challenges. Therefore, using a more holistic combination of methods could offer a viable solution. Future research should consider larger sample sizes and multicenter external evaluations to address these limitations.

Conclusions

In conclusion, we have successfully translated, culturally adapted, and validated DigiHealthCom for Chinese health care professionals. Our findings demonstrate that the Chinese version of DigiHealthCom is a reliable and valid instrument for assessing digital health competence among these health care professionals.

Acknowledgments

The authors would like to acknowledge all the participants who participated in this study; in particular, Dr Donglan Ling for help with the translation process. XL is supported by the National Natural Science Foundation of China (grant 82171202). The funding agents had no role in the study design; collection, analysis, and interpretation of the data; writing of the manuscript; or decision to submit the paper for publication.

Data Availability

The dataset collected in this study is available from the corresponding author on reasonable request.

Authors' Contributions

XL and LG had full access to all the data in this study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Concept and design were contributed by XL. Acquisition, analysis, or interpretation of data were contributed by LG, MC, and JW. Drafting of the manuscript was handled LG and XL. Critical revision of the manuscript for important intellectual content were contributed by XL, LG, JW, MC, and JW. Administrative, technical, or material support were contributed by JW. The authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Modified items of the Chinese version of the DigiHealthCom (Digital Health Competence) instrument after cultural adaption.
[PDF File, 132 KB - humanfactors_v12i1e65373_app1.pdf]

Multimedia Appendix 2
Comparison of DigiHealthCom (Digital Health Competence) scores between health care professionals in the Southern China group and in the Northern and Western China group.
[DOCX File, 16 KB - humanfactors_v12i1e65373_app2.docx]

Checklist 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[XLSX File, 14 KB - humanfactors_v12i1e65373_app3.xlsx]

References

1. World Health Organization. Global strategy on digital health 2020-2025. 2021. URL: https://www.who.int/publications/i/item/9789240020924 [accessed 2024-08-04]
2. World Health Organization. WHO guideline recommendations on digital interventions for health system strengthening. 2019. URL: https://iris.who.int/handle/10665/311980 [accessed 2024-08-04]
3. Qin S, Zhang J, Sun X, et al. A scale for measuring nursing digital application skills: a development and psychometric testing study. BMC Nurs 2024 May 31;23(1):366. [doi: 10.1186/s12912-024-02030-8] [Medline: 38822276]
4. Grand View Research. Digital health market size, share & trends analysis report by technology (healthcare analytics, mhealth), by component (hardware, software, services), by application, by end-use, by region, and segment forecasts, 2024 - 2030. URL: https://www.grandviewresearch.com/industry-analysis/digital-health-market [accessed 2025-02-14]
5. Nazeha N, Pavagadhi D, Kyaw BM, Car J, Jimenez G, Tudor Car L. A digitally competent health workforce: scoping review of educational frameworks. J Med Internet Res 2020 Nov 5;22(11):e22706. [doi: 10.2196/22706] [Medline: 33151152]
6. Odendaal WA, Anstey Watkins J, Leon N, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942. [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
7. Bichel-Findlay J, Koch S, Mantas J, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics: Second Revision. Int J Med Inform 2023 Feb;170(104908):104908. [doi: 10.1016/j.ijmedinf.2022.104908] [Medline: 36502741]
8. Jarva E, Oikarinen A, Andersson J, et al. Healthcare professionals' perceptions of digital health competence: a qualitative descriptive study. Nurs Open 2022 Mar;9(2):1379-1393. [doi: 10.1002/nop2.1184] [Medline: 35094493]
9. Konttila J, Siira H, Kyngäs H, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]
10. Brown J, Pope N, Bosco AM, Mason J, Morgan A. Issues affecting nurses' capability to use digital technology at work: an integrative review. J Clin Nurs 2020 Aug;29(15-16):2801-2819. [doi: 10.1111/jocn.15321] [Medline: 32416029]
11. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252. [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
12. Salahuddin L, Ismail Z. Classification of antecedents towards safety use of health information technology: a systematic review. Int J Med Inform 2015 Nov;84(11):877-891. [doi: 10.1016/j.ijmedinf.2015.07.004] [Medline: 26238706]
13. Wernhart A, Gahbauer S, Haluza D. eHealth and telemedicine: Practices and beliefs among healthcare professionals and medical students at a medical university. PLoS ONE 2019;14(2):e0213067. [doi: 10.1371/journal.pone.0213067] [Medline: 30818348]
14. Alam K, Mahumud RA, Alam F, Keramat SA, Erdiaw-Kwasie MO, Sarker AR. Determinants of access to eHealth services in regional Australia. Int J Med Inform 2019 Nov;131(103960):103960. [doi: 10.1016/j.ijmedinf.2019.103960] [Medline: 31518858]

15. Longhini J, Rossettini G, Palese A. Digital health competencies among health care professionals: systematic review. J Med Internet Res 2022 Aug 18;24(8):e36414. [doi: 10.2196/36414] [Medline: 35980735]
16. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27. [doi: 10.2196/jmir.6709] [Medline: 28119275]
17. Kayser L, Karnoe A, Furstrand D, et al. A multidimensional tool based on the eHealth Literacy Framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36. [doi: 10.2196/jmir.8371] [Medline: 29434011]
18. Cowan DT, Norman I, Coopamah VP. Competence in nursing practice: a controversial concept--a focused review of literature. Nurse Educ Today 2005 Jul;25(5):355-362. [doi: 10.1016/j.nedt.2005.03.002] [Medline: 15904996]
19. Jarva E, Oikarinen A, Andersson J, Tomietto M, Kääriäinen M, Mikkonen K. Healthcare professionals' digital health competence and its core factors; development and psychometric testing of two instruments. Int J Med Inform 2023 Mar;171(104995):104995. [doi: 10.1016/j.ijmedinf.2023.104995] [Medline: 36689840]
20. Jarva E, Oikarinen A, Andersson J, Pramila-Savukoski S, Hammarén M, Mikkonen K. Healthcare professionals' digital health competence profiles and associated factors: a cross-sectional study. J Adv Nurs 2024 Aug;80(8):3236-3252. [doi: 10.1111/jan.16096] [Medline: 38323687]
21. Cruchinho P, López-Franco MD, Capelas ML, et al. Translation, cross-cultural adaptation, and validation of measurement instruments: a practical guideline for novice researchers. J Multidiscip Healthc 2024;17(2701-2728):2701-2728. [doi: 10.2147/JMDH.S419714] [Medline: 38840704]
22. Wenjuanxing. wjx. URL: https://www.wjx.cn [accessed 2024-11-07]
23. Zhang C, Yang Z, Zhang H. Psychometric evaluation of the Chinese version of occupational lowback pain prevention behaviors questionnaire among clinical nurses: a validation study. Front Public Health 2022;10(827604):827604. [doi: 10.3389/fpubh.2022.827604] [Medline: 35400039]
24. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
25. Jalali A, Naghibzadeh A, Mohammadi MM, et al. Translation and validation of the Persian version of the Nursing Practice Readiness Scale (NPRS) for new graduate nurses. BMC Nurs 2024 Oct 16;23(1):760. [doi: 10.1186/s12912-024-02434-6] [Medline: 39415162]
26. El-Ammari A, El Malki H, Moutawakkil SG, et al. Validation of the Center for Epidemiologic Studies Depression Scale (CES-D) in a Moroccan sample with substance use disorder. BMC Psychiatry 2023 Oct 6;23(1):723. [doi: 10.1186/s12888-023-05245-2] [Medline: 37803359]
27. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34. [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
28. Alelayan H, Liang L, Ye R, Aldosari N, Liao X. Translation and linguistic validation of the DISABKIDS chronic generic module into simplified Chinese (DCGM-37) for use among children with cancer. J Spec Pediatr Nurs 2022 Jul;27(3):e12374. [doi: 10.1111/jspn.12374] [Medline: 35415867]
29. Kaihlaniemi J, Liljamo P, Rajala M, Kaakinen P, Oikarinen A. Health care professionals' experiences of counselling competence in digital care pathways - a descriptive qualitative study. Nurs Open 2023 Jul;10(7):4773-4785. [doi: 10.1002/nop2.1729] [Medline: 36960773]
30. Nittari G, Khuman R, Baldoni S, et al. Telemedicine practice: review of the current ethical and legal challenges. Telemed J E Health 2020 Dec;26(12):1427-1437. [doi: 10.1089/tmj.2019.0158] [Medline: 32049608]
31. Kaihlanen AM, Elovainio M, Virtanen L, et al. Nursing informatics competence profiles and perceptions of health information system usefulness among registered nurses: A latent profile analysis. J Adv Nurs 2023 Oct;79(10):4022-4033. [doi: 10.1111/jan.15718] [Medline: 37243421]
32. Kaihlanen AM, Virtanen L, Buchert U, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188. [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
33. Ahonen O, Kouri P, Salanterä S, Liljamo P, Kinnunen UM, Saranto K, et al. Finnish nurses association's digital social and health services strategy. Finnish Nurses Association. URL: https://iris.who.int/handle/10665/366710 [accessed 2024-11-05]
34. Tynkkynen LK, Karanikolos M, Litvinova Y. Finland: Health system summary. 2023. URL: https://iris.who.int/handle/10665/366710
35. Wu G, Gong M, Wu Y, Liu L, Shi B, Zeng Z. Advancing digital health in China: aligning challenges, opportunities, and solutions with the Global Initiative on Digital Health (GIDH). Health Care Sci 2024 Oct;3(5):365-369. [doi: 10.1002/hcs2.118] [Medline: 39479272]
36. Yu-Tong T, Yan Z, Zhen L, Bing X, Qing-Yun C. Telehealth readiness and its influencing factors among Chinese clinical nurses: a cross-sectional study. Nurse Educ Pract 2022 Jan;58(103278):103278. [doi: 10.1016/j.nepr.2021.103278] [Medline: 34954659]

Abbreviations

AVE: average variance extracted
CFA: Confirmatory Factor Analysis
CFI: Comparative Fit Index
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CR: composite reliability
DHLI: Digital Health Literacy Instrument
DigiHealthCom: Digital Health Competence
eHLQ: eHealth literacy questionnaire
HTMT: heterotrait-monotrait ratio
I-CVI: item content validity index
ICC: Intraclass correlation coefficient
ICT: information and communication technology
IFI: Incremental Fit Index
NDASS: Nursing Digital Application Skill Scale
RSMEA: root-mean-square error of approximation
S-CVI/Ave: Scale CVI/Average
SRMR: standardized root-mean-square residual
TLI: Tucker-Lewis Index
WHO: World Health Organization

Edited by A Kushniruk; submitted 14.08.24; peer-reviewed by A Mainz, E Jarva, YC Uscanga; revised version received 03.02.25; accepted 10.02.25; published 21.03.25.

Please cite as:
Gao L, Chen M, Wei J, Wang J, Liao X
The Chinese Version of the DigiHealthCom (Digital Health Competence) Instrument for Assessing Digital Health Competence of Health Care Professionals: Translation, Adaptation, and Validation Study
JMIR Hum Factors 2025;12:e65373
URL: https://humanfactors.jmir.org/2025/1/e65373
doi:10.2196/65373

Application of Clinical Department–Specific AI-Assisted Coding Using Taiwan Diagnosis-Related Groups: Retrospective Validation Study

An-Tai Lu[1,2], MSc; Chong-Sin Liou[3], MSc; Chia-Hsin Lai[3]; Bo-Tsz Shian[3]; Ming-Ta Li[2], BBA; Chih-Yen Sun[4], BSc; Hao-Yun Kao[1*], PhD; Hong-Jie Dai[5,6,7,8*], PhD; Ming-Ju Tsai[2,9,10*], PhD

[1]Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, No.100, Shih-Chuan 1st Road, Sanmin Dist, Kaohsiung, Taiwan

[10]Department of Internal Medicine, School of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[2]Department of Medical Records, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[3]Intelligent System Lab, College of Electrical Engineering and Computer Science, Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[4]Department of Information Technology, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[5]Department of Electrical Engineering, National Kaohsiung University of Science and Technology, Kaohsiung, Taiwan

[6]School of Post-Baccalaureate Medicine, Kaohsiung Medical University, Kaohsiung, Taiwan

[7]National Institute of Cancer Research, National Health Research Institutes, Tainan, Taiwan

[8]Center for Big Data Research, Kaohsiung Medical University, Kaohsiung, Taiwan

[9]Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan

[*]these authors contributed equally

Corresponding Author:
Hao-Yun Kao, PhD
Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, No.100, Shih-Chuan 1st Road, Sanmin Dist, Kaohsiung, Taiwan

Abstract

Background: The accuracy of the *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) procedure coding system (PCS) is crucial for generating correct Taiwan diagnosis-related groups (DRGs), as coding errors can lead to financial losses for hospitals.

Objective: The study aimed to determine the consistency between an artificial intelligence (AI)-assisted coding module and manual coding, as well as to identify clinical specialties suitable for implementing the developed AI-assisted coding module.

Methods: This study examined the AI-assisted coding module from the perspective of health care professionals. The research period started in February 2023. The study excluded cases outside of Taiwan DRGs, those with incomplete medical records, and cases with Taiwan DRG disposals *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision)* PCS. Data collection was conducted through retrospective medical record review. The AI-assisted module was constructed using a hierarchical attention network. The verification of the Taiwan DRGs results from the AI-assisted coding model focused on the major diagnostic categories (MDCs). Statistical computations were conducted using SPSS version 19. Research variables consisted of categorical variables represented by MDC, and continuous variables were represented by the relative weight of Taiwan DRGs.

Results: A total of 2632 discharge records meeting the research criteria were collected from February to April 2023. In terms of inferential statistics, κ statistics were used for MDC analysis. The infectious and parasitic diseases MDC, as well as the respiratory diseases MDC had κ values exceeding 0.8. Clinical inpatient specialties were statistically analyzed using the Wilcoxon signed rank test. There was not a difference in coding results between the 23 clinical departments, such as the Division of Cardiology, the Division of Nephrology, and the Department of Urology.

Conclusions: For human coders, with the assistance of the *ICD-10-CM* AI-assisted coding system, work time is reduced. Additionally, strengthening knowledge in clinical documentation enables human coders to maximize their role. This positions them to become clinical documentation experts, preparing them for further career development. Future research will apply the same method to validate the *ICD-10* AI-assisted coding module.

(JMIR Hum Factors 2025;12:e59961) doi:10.2196/59961

KEYWORDS

diagnosis-related group; artificial intelligence coding; International Classification of Diseases, Tenth Revision, Clinical Modification; ICD-10-CM; coding professionals

Introduction

The *International Statistical Classification of Diseases* (*ICD*) system was set up by the World Health Organization (WHO) for the purpose of tracking diseases globally. Over the past several decades, the WHO has made significant changes to both content and structure. It accompanies a new scientific understanding of diseases and new structures for organizing *ICD* codes to accommodate enhanced use and extensibility [1]. The WHO introduced the *ICD* in 1948, and it is a universal language used to categorize diseases or causes of death. The use of it is attributed to health care–related units in 194 countries and is generated by professional coders based on discharge records, with countries adjusting the *ICD* to their circumstances. In 2016, Taiwan adopted the international trend of switching from *ICD-9-CM* (*International Classification of Diseases, Ninth Revision, Clinical Modification*) to *ICD-10-CM* (*International Classification of Diseases, Tenth Revision, Clinical Modification*) procedure coding system (PCS) for coding hospital patient diagnoses, procedures, analysis and reimbursement. The National Health Insurance Administration (NHIA) under the Ministry of Health and Welfare has adopted the 2014 edition of *ICD-10-CM* PCS, with approximately 71,900 diagnosis codes and 78,500 procedure codes.

The use of the *ICD-10-CM* PCS involves coding and classifying morbidity data from health records, reimbursement claims, and administrative databases. Improving health care quality, monitoring public health, and conducting research are all benefits of the *ICD-10-CM* PCS in Taiwan and involves converting the physician's discharge diagnosis into *ICD-10-CM* codes by following the primary diagnosis selection principle announced by the NHIA. The diagnosis-related group (DRG) provides information such as health insurance reimbursement, relative weight, presence of comorbidities, and complications for the current hospitalization.

The accuracy of *ICD-10-CM* PCS coding is crucial for generating accurate Taiwan DRGs, as coding errors can lead to financial losses for hospitals [2,3]. According to the coding principles set forth by the NHIA and the Taiwan Society of Medical Records Management, coding is based on the inpatient and emergency room records of patients. In the past, this task was undertaken by trained and certified clinical coding professionals (referred to as coding professionals hereafter), but with the rapid advances of medical technology, the rules of disease classification have also evolved, and coding professionals must regularly accumulate relevant training hours to update their disease classification skills [4].

In recent years, artificial intelligence (AI) and natural language processing have shown exciting potential in the field of automatic clinical coding. In 2021, the disease coding scales in the United States were worth approximately 18 billion US dollars. Several technology companies in the United States have developed AI-assisted coding systems, and scholars believe that interdisciplinary collaboration and feedback from clinical coding professionals are essential to further refine the modules [5,6]. Research on AI-assisted coding consistently conclude that it improves quality and reduces error rates while saving costs [7,8]. AI-assisted *ICD-10-CM* PCS coding can be considered as a text classification task within the field of machine learning [9]. In recent years, studies in the machine learning text classification field have predominantly proposed using deep learning–based neural networks [10]. Many research papers have focused on AI assistance in *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision)* coding [11-14], but few have examined the results of coding implementation from the perspective of disease classification personnel. The development and validation process of the AI-assisted coding model requires the involvement and feedback of clinical coders to enhance accuracy and correctness, aligning with user needs [15].

In Taiwan, several hospitals have also ventured into the development of AI-assisted coding for disease classification. However, due to variations in physicians' documentation of medical records across different hospitals, the AI-assisted coding systems developed are not universally applicable [11], necessitating the development and validation of customized AI-assisted coding systems. Medical coding personnel must review the discharge records meticulously and then translate the discharge diagnoses and procedures (interventions) recorded in the medical records into *ICD-10* codes. In the past, the most significant factor contributing to coding errors was handwritten medical records by physicians, which were difficult to decipher or included abbreviations, leading to mistakes [16]. In recent years, most medical centers in Taiwan have adopted electronic health records, resulting in a significant reduction in coding errors caused by handwritten records. Clinical coding personnel also encounter various pressures, including the need to accomplish all inpatient coding tasks within specified deadlines, optimize Taiwan DRGs assignment coding, enhance and maintain coding reliability and validity, and engage in discussions with clinical physicians regarding the content of medical record writing.

Recently, the global trend in AI coding has been on the rise [11,13,17-19]. In this study, we have developed an exclusive *ICD-10-CM* AI-assisted coding module. Coding professionals took part in the research and offered suggestions to improve the efficiency of coding operations. Consequently, this study focuses primarily on the following two research aims: (1) to verify the consistency between the AI-assistant coding module and a coding professional in encoding, based on the MDC results in Taiwan DRGs and (2) to find the clinical departments within the medical center that can benefit from using the developed AI-assisted coding module.

Methods

Data Description

This study used a total of 136,841 unstructured discharge summaries of patients who were hospitalized, recorded in Kaohsiung Medical University Chung-Ho Memorial Hospital from April 1, 2019, to December 31, 2020, as the primary data source. Textbox 1 displays an example of a discharge summary from the Kaohsiung Medical University Chung-Ho Memorial Hospital.

Textbox 1. Example discharge summary.

Chief complaint:

Abdominal pain for 1 day

Impression on admission:

Sepsis, focus on retroperitoneal abscess

Discharge diagnosis:

Sepsis, focus on retroperitoneal abscess due to surgical site infection

History on admission:

This time, according to the patient's statement, he suffered from recurrent abdominal dull pain after discharge. The pain was serious by jejunostomy feeding, and there was no relieving factor. The pain suddenly progressed...

This study verified the AI-assisted coding from the perspective of coding professionals. Since the AI-assisted coding system was introduced in the medical center in February 2023, the study period began in February 2023. The subjects of this study were selected based on the following exclusion criteria: non–Taiwan DRGs cases, cases with procedures (*ICD-10* PCS) in Taiwan DRGs, and cases with incomplete medical records. According to the study conditions, there were approximately 700 to 1000 cases per month. The coding by both the AI-assisted coding module and coding professionals were based on the electronic discharge summaries of a certain medical center each month.

Research Design

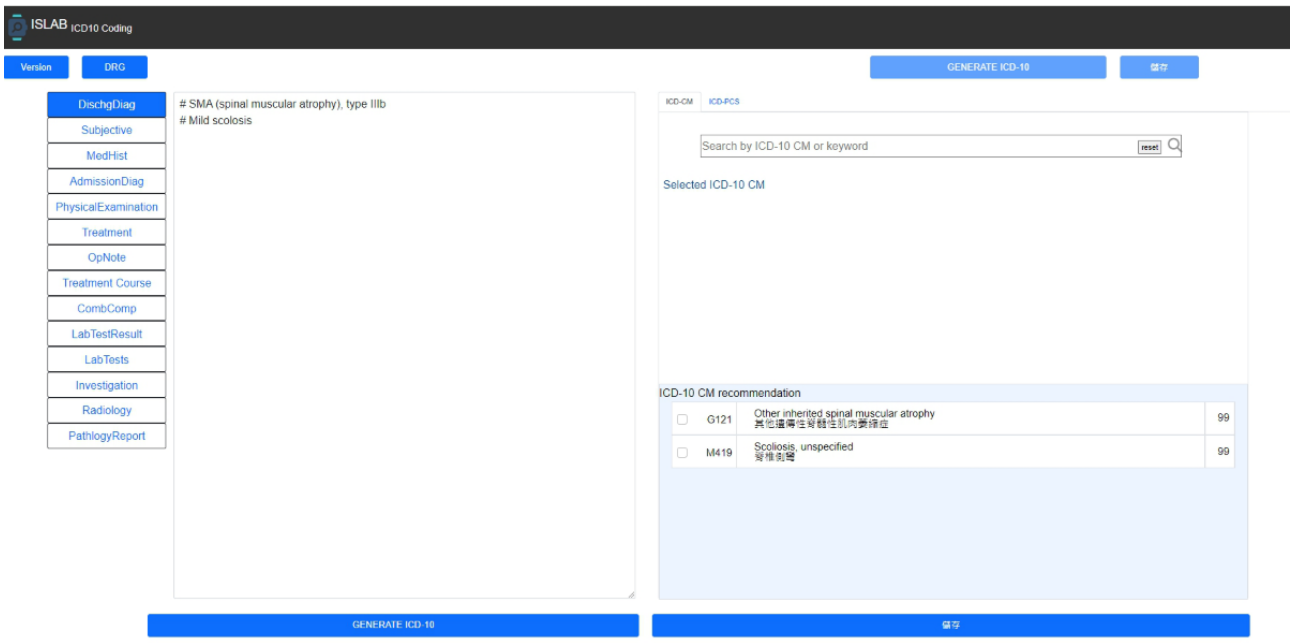
After each data entry was encoded by the AI-assisted coding module and verified by a coding professional, it was transmitted to a certain university's database. The results of both the AI-assisted coding and coding professional were compared using an Excel (Microsoft) file. Following the linkage to the NHIA's DRG calculation software, separate datasets for Taiwan DRGs were obtained for both the AI-assisted coding and coding personnel, with the consistency of the primary diagnosis coding between these two groups being examined. In cases of discrepancies, the medical records were scrutinized again by the coding professional to determine if the AI-assisted coding results met the criteria for primary coding as per the coding professional; the consistency results of the Taiwan DRGs data for both the AI-assisted coding and coding professional were adjusted accordingly.

AI-Assisted Coding Construction Process

The AI-assisted *ICD-10-CM* coding system was developed by CSL, CHL, and BTS, based on approximately 110,000 discharge summaries collected from April 1, 2019 to December 31, 2020, in a medical center. The deidentified summary data were processed by segmenting sentences and filtering out meaningless delimiters and prefix symbols (eg, # or "") by using a clinical natural language processing tool [20]. The data were categorized into 21 groups based on the first 3 codes of the *ICD-10-CM*, and models were built using bidirectional encoder representations from transformers (BERTs) [21] and hierarchical attention networks (HANs) [22]. The results favored HANs, leading to the decision to adopt the HAN module. The precision, recall, and F1 scores of the developed HAN model were 0.55, 0.82, and 0.66, respectively. For the top 50 most frequent codes, the F1 score of the developed HAN model was 0.818.

Aside from module modeling, another time-consuming task was the design of the user interface for the coding professionals, as it needed to present discharge summaries, laboratory data, and imaging reports, as well as the *ICD-10-CM* codes predicted by the AI-assisted coding module. Coding professionals were actively involved in providing feedback during the interface design process. Figure 1 provides an illustration of the designed user interface, which provided suggestions automatic *ICD-10-CM* recommendation and fields for coding professionals to input the final codes. The developed AI-coding system was integrated into a medical center's hospital information systems in November 2022 and operated in February 2023.

Figure 1. User interface screenshot.



Ethical Considerations

This study was approved by Kaohsiung Medical University Ching-Ho Memorial Hospital (institutional review boards number: KMUHIRB-E(II)-20230214). The institutional review board approval covered secondary analysis without additional consent. Data was anonymized or deidentified. There was not any compensation provided to participants.

Statistical Analysis

The study involved an analysis incorporating descriptive statistics for exploration, as well as inferential statistics for investigating MDCs and relative weight. Statistical computations were conducted using SPSS version 19. Research variables consisted of categorical variables represented by MDC, and continuous variables were represented by the relative weight of Taiwan DRGs.

Results

Distribution of *ICD-10* Codes

The distribution of the *ICD-10* codes seen in the collected training dataset is shown in Multimedia Appendix 1. The first digit of the *ICD-10-CM* code consisted of English letters, so the alphabetical characters on the horizontal axis of the log data were the first digit of the *ICD-10-CM* code, showing diseases pertaining to different systems. According to Multimedia Appendix 1, data starting with codes C, E, and I in *ICD-10-CM* had the highest volume, with C representing neoplastic diseases; E for endocrinal, nutritional, and metabolic diseases; and I for diseases of the circulatory system. These were the body systems with the highest learning volumes for the AI-assisted coding module.

Descriptive Statistics

In the period from February to April 2023, a total of 15,756 discharges were recorded. Excluding cases with interventions, non–Taiwan DRG cases, and cases with incomplete medical records, there was a total of 2632 cases. The primary diagnosis was the key factor in deciding the main disease category, while secondary diagnoses only affected the distribution of Taiwan DRGs within the same primary disease category. According to disease classification rules, the primary diagnosis was based on the reason for the patient's admission, but only one disease could be selected as the primary diagnosis. If multiple diseases were treated during admission, selecting any one of them as the primary diagnosis was not considered an error. Therefore, the coding professional (author ATL) manually examined the discharged cases' notes to categorize the output of the AI-assisted system into one of the following categories. The results are shown in Table 1.

Table . Frequency distribution and percentage analysis of primary diagnoses.

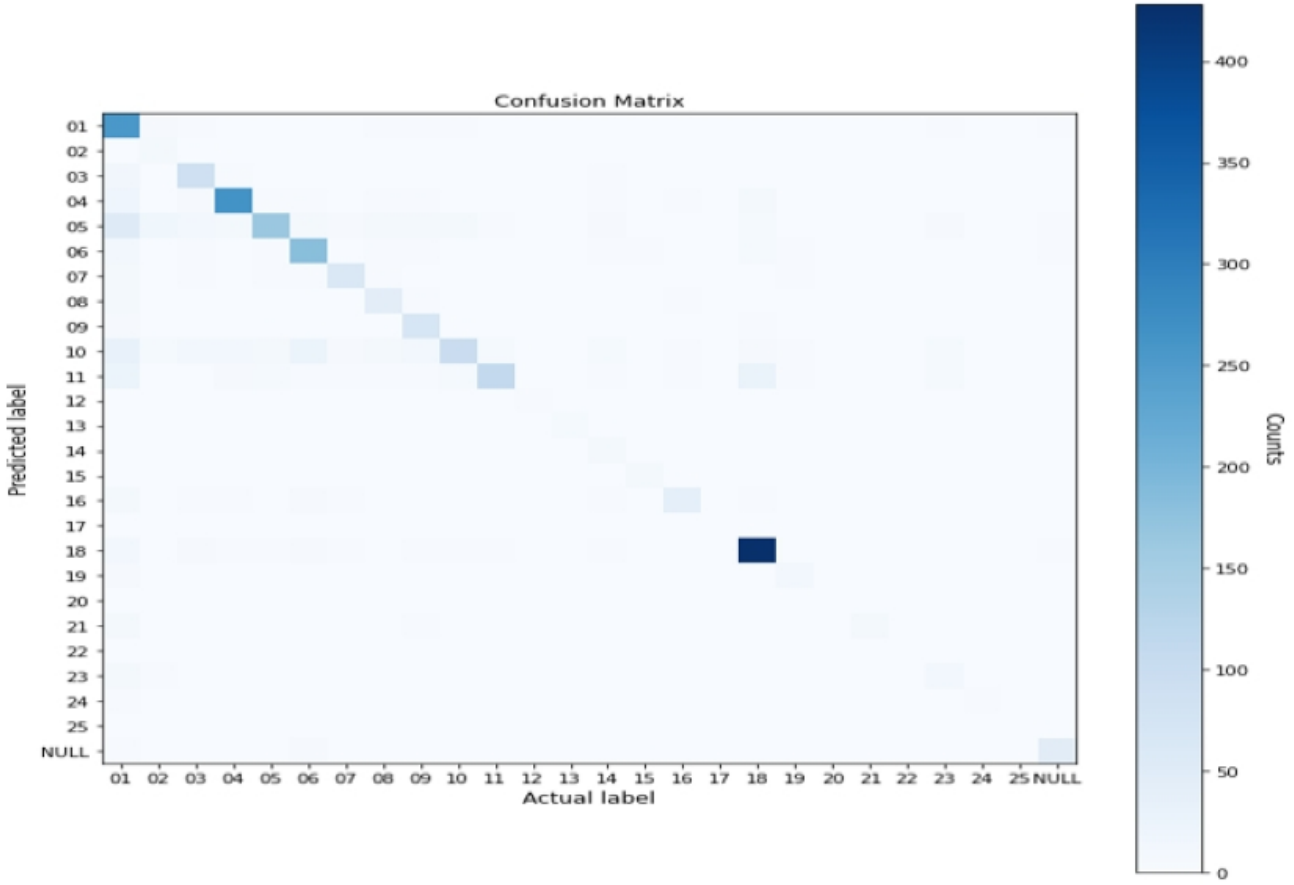
Variable	Month of case, n (%)		
	February (n=748)	March (n=991)	April (n=893)
No primary diagnosis	181 (24.2)	277 (28)	285 (31.9)
Incorrect secondary diagnosis with a primary diagnosis	462 (61.8)	477 (48.1)	369 (41.3)
All correct	79 (10.6)	181 (18.3)	177 (19.8)
All incorrect	26 (3.5)	56 (5.7)	62 (7)

Operational definitions were as follows:

- No primary diagnosis: in comparison to the coding professional, a single hospitalization's predicted diagnosis codes did not include a primary diagnosis.
- Incorrect secondary diagnosis with a primary diagnosis: in comparison to the coding professional, a single hospitalization's predicted diagnosis codes included a primary diagnosis, but there was at least 1 error in the secondary diagnoses.
- All correct: all predicted diagnosis codes for a single hospitalization perfectly aligned with those given by the coding professional.
- All incorrect: in comparison to the disease classification personnel, none of the predicted diagnosis codes in a single hospitalization were the same.

In Figure 2, we analyzed the agreement of MDC classification between the AI-assisted coding module and the coding professional through a heat map analysis. The vertical and horizontal axes in Figure 2 represent MDCs coded by the AI-assisted coding module and MDCs coded by coding professionals, respectively. The intensity of color in the figure indicated a higher number of agreed MDCs between the AI-assisted coding module and professionals. As shown in Figure 2, MDC 1 (diseases and disorders of the nervous system), MDC 4 (diseases and disorders of the respiratory system), and MDC 18 (infectious and parasitic diseases and disorders) had the highest agreements.

Figure 2. MDC heat map analysis between AI-coding module and professionals.



The κ Coefficient Test

Furthermore, we assessed the MDC agreement between the AI coding module and coding professionals using the κ coefficient test. The κ values were broadly categorized into 5 groups based on various levels of agreement: extremely low agreement (0.00-0.20), fair agreement (0.21-0.40), moderate agreement (0.41-0.60), high agreement (0.61-0.80), and almost perfect agreement (0.81-1.0).

When analyzing the cumulative data for February to April 2023 (Table 2), the MDCs with the highest consistency were MDC 4 (diseases and disorders of the respiratory system) and MDC 18 (infectious and parasitic diseases and disorders), followed by MDC 1 (diseases and disorders of the nervous system), MDC 3 (diseases and disorders of the ear, nose, mouth and throat), MDC 6 (diseases and disorders of the digestive system), MDC 7 (diseases and disorders of the hepatobiliary system and pancreas), MDC 9 (diseases and disorders of the skin, subcutaneous tissue and breast), MDC 11 (diseases and disorders of the kidney and urinary tract), MDC 13 (diseases and disorders of the female reproductive system), MDC 15 (newborn and other neonates), and MDC 16 (diseases and disorders of the blood and blood forming organs and immunological disorders).

Table . Kappa tests for aggregation of major diagnostic category in the total counts for February to April 2023 (κ=0.592).

No.	Major diagnostic category	AI-assisted case coding (n=2362), n (%)	Cases coded by human coders (n=2362), n (%)	Kappa value
1	Diseases and disorders of the nervous system	280 (10.6)	509 (19.3)	0.670[a]
2	Diseases and disorders of the eye	9 (0.3)	38 (1.4)	0.300
3	Diseases and disorders of the ear, nose, mouth, and throat	113 (4.3)	132 (5)	0.689[a]
4	Diseases and disorders of the respiratory system	309 (11.7)	302 (11.5)	0.845[b]
5	Diseases and disorders of the circulatory system	310 (11.8)	184 (7)	0.607
6	Diseases and disorders of the digestive system	217 (8.2)	229 (8.7)	0.775[a]
7	Diseases and disorders of the hepatobiliary system and pancreas	90 (3.4)	84 (3.2)	0.710[a]
8	Diseases and disorders of the musculoskeletal system and connective tissue	66 (2.5)	87 (3.3)	0.576
9	Diseases and disorders of the skin, subcutaneous tissue, and breast	83 (3.2)	116 (4.4)	0.692[a]
10	Diseases and disorders of the endocrine, nutritional, and metabolic systems	237 (9)	132 (5)	0.505
11	Diseases and disorders of the kidney and urinary tract	205 (7.8)	120 (4.6)	0.648[a]
12	Diseases and disorders of the male reproductive system	7 (0.3)	3 (0.1)	0.362
13	Diseases and disorders of the female reproductive system	8 (0.3)	8 (0.3)	0.749[a]
14	Pregnancy, childbirth and puerperium	11 (0.4)	41 (1.6)	0.419
15	Newborn and other neonates (perinatal period)	9 (0.3)	15 (0.6)	0.635[a]
16	Diseases and disorders of the blood and blood forming organs and immunological disorders	64 (2.4)	57 (2.2)	0.624[a]
17	Myeloproliferative diseases and disorders (poorly differentiated neoplasms)	3 (0.1)	2 (0.1)	0.399
18	Infectious and parasitic diseases and disorders	465 (17.7)	505 (19.2)	0.870[b]
19	Mental diseases and disorders	23 (0.9)	0 (0)	—[c]
20	Alcohol or drug abuse or induced mental disorder	2 (0.1)	0 (0)	—
21	Injuries, poison, and toxic effects of drugs	24 (0.9)	21 (0.8)	0.404
22	Burns	0 (0)	1 (0)	—

No.	Major diagnostic category	AI-assisted case coding (n=2362), n (%)	Cases coded by human coders (n=2362), n (%)	Kappa value
23	Factors influencing health status and other contacts with health services	24 (0.9)	42 (1.6)	0.366
24	Multiple significant trauma	6 (0.2)	4 (0.2)	0.599
25	HIV infection	2 (0.1)	0 (0)	—
—	None	65 (2.5)	0 (0)	—

[a]High agreement (0.61-0.80).

[b]Almost perfect agreement (0.81-1.00).

[c]Kappa value was not calculated when there were 0 cases in a coding group.

Inferential Statistical Analysis: Wilcoxon Signed Rank Test

The κ coefficient test was used for a broad-scale analysis of MDCs. However, under the same MDC, it was possible to further classify the data into numerous Taiwan DRGs, with each having its own code and relative weight. Even within the same MDC, this might result in different Taiwan DRGs. Furthermore, some diseases could be treated across departments. Therefore, for the statistical analysis of relative weight, we first conducted a normality analysis of the relative weights obtained from both AI-assisted coding and coding professionals. The statistical results based on the Kolmogorov-Smirnov analysis yielded a significance level of less than .05, showing a nonnormal distribution. Given that the research sample consisted of paired data, the nonparametric Wilcoxon signed rank test was used to analyze whether there were differences in relative weight between AI-assisted coding and coder-assigned coding; the null hypothesis assumed that there was no difference in relative weight between AI-assisted coding and coder-assigned coding.

The Wilcoxon signed rank test, with clinical departments as the unit of analysis, identified differences in relative weight in the following 12 departments: Division of Endocrinology and Metabolism, Division of Hematology and Oncology, Division of General Internal Medicine, Division of Geriatrics and Gerontology, Division of Trauma, Division of Neurosurgery, Division of Cardiovascular Surgery, Division of General and Digestive Surgery, Division of Pediatric Neurology, Department of Otorhinolaryngology, Department of Neurology, and Department of Rehabilitation Medicine. As shown in Table 3, the overall statistical result with a *P* value of <.001 showed that there were still differences between AI-assisted coding and coder-assigned coding in this study.

Table . Wilcoxon signed rank test results across various clinical departments (P<.001).

Clinical department	Frequency of cases (n=2632), n (%)	Relative weight (95%CI)		*P* value
		AI[a] coding	Human coding	
Division of Gastroenterology	61 (2.3)	0.66 (0.58 - 0.74)	0.70 (0.63 - 0.77)	.12
Division of Hepatobiliary and Pancreatic Medicine	81 (3.1)	0.69 (0.63 - 0.75)	0.74 (0.69 - 0.79)	.07
Division of Cardiology	177 (6.7)	0.69 (0.65 - 0.73)	0.72 (0.68 - 0.75)	.39
Division of Chest Medicine	181 (6.9)	0.90 (0.85 - 0.93)	0.90 (0.87 - 0.94)	.50
Division of Nephrology	64 (2.4)	0.71 (0.65 - 0.78)	0.71 (0.64 - 0.78)	.91
Division of Endocrinology and Metabolism	36 (1.4)	0.62 (0.54 - 0.71)	0.68 (0.59 - 0.77)	.03
Division of Hematology and Oncology	42 (1.6)	0.79 (0.69 - 0.88)	0.87 (0.80 - 0.94)	.006
Division of Rheumatology, Immunology, and Allergology	36 (1.4)	0.68 (0.58 - 0.79)	0.71 (0.64 - 0.79)	.33
Division of Infectious Diseases	78 (3)	0.91 (0.85 - 0.98)	0.94 (0.90 - 0.99)	.39
Division of General Internal Medicine	205 (7.8)	0.69 (0.65 - 0.73)	0.73 (0.69 - 0.77)	.005
Division of Geriatrics and Gerontology	54 (2.1)	0.92 (0.85 - 0.99)	0.99 (0.94 - 1.04)	.003
Division of Trauma	16 (0.6)	0.43 (0.21 - 0.65)	0.64 (0.46 - 0.83)	.04
Division of Neurosurgery	151 (5.7)	0.53 (0.48 - 0.58)	0.68 (0.63 - 0.73)	<.001
Division of Cardiovascular Surgery	24 (0.9)	0.67 (0.54 - 0.80)	0.94 (0.82 - 1.06)	.002
Division of Chest Surgery	14 (0.5)	0.56 (0.43 - 0.69)	0.60 (0.45 - 0.75)	.40
Division of Pediatric Surgery	9 (0.3)	0.45 (0.31 - 0.59)	0.48 (0.33 - 0.63)	.28
Division of Plastic Surgery	9 (0.3)	0.64 (0.38 - 0.90)	0.68 (0.44 - 0.92)	.89
Division of Colorectal Surgery	46 (1.7)	0.54 (0.45 - 0.62)	0.59 (0.52 - 0.66)	.16
Division of Breast Oncology and Surgery	16 (0.6)	0.47 (0.32 - 0.62)	0.62 (0.48 - 0.75)	.05
Division of General and Digestive Surgery	56 (2.1)	0.56 (0.49 - 0.62)	0.63 (0.57 - 0.68)	.009
Department of Gynecology Obstetrics	60 (2.3)	0.47 (0.39 - 0.55)	0.46 (0.40 - 0.52)	.72
Division of Pediatric Hematology and Oncology	42 (1.6)	0.45 (0.38 - 0.52)	0.47 (0.40 - 0.55)	.44
Division of Pediatric Cardiology and Pulmonology	86 (3.3)	0.52 (0.37 - 0.66)	0.53 (0.43 - 0.63)	.24
Division of Pediatric Neurology	92 (3.5)	0.43 (0.34 - 0.52)	0.45 (0.40 - 0.50)	.008
Division of Neonatology	14 (0.5)	0.87 (0.28 - 1.47)	0.81 (0.43 - 1.18)	.69
Division of General Pediatrics	299 (11.4)	0.39 (0.36 - 0.41)	0.39 (0.37 - 0.41)	.56
Division of Pediatric Allergy Immunology	8 (0.3)	0.31 (0.17 - 0.46)	0.35 (0.25 - 0.46)	.32

Clinical department	Frequency of cases (n=2632), n (%)	Relative weight (95%CI)		*P* value
		AI[a] coding	Human coding	
Department of Otorhinolaryngology	53 (2)	0.57 (0.50 - 0.63)	0.51 (0.45 - 0.58)	.02
Ophthalmology Department	12 (0.5)	0.42 (0.34 - 0.50)	0.44 (0.36 - 0.52)	.44
Department of Orthopaedics	13 (0.5)	0.40 (0.25 - 0.55)	0.52 (0.39 - 0.65)	.16
Department of Urology	46 (1.7)	0.63 (0.56 - 0.70)	0.62 (0.55 - 0.69)	.95
Department of Dermatology	87 (3.3)	0.42 (0.37 - 0.48)	0.41 (0.35 - 0.46)	.18
Department of Neurology	366 (13.9)	0.66 (0.63 - 0.69)	0.72 (0.69 - 0.75)	<.001
Division of Family Medicine	49 (1.9)	0.96 (0.87 - 1.04)	1.00 (0.93 - 1.07)	.17
Department of Rehabilitation Medicine	47 (1.8)	0.85 (0.75 - 0.94)	1.21 (1.14 - 1.28)	<.001
Department of Psychiatry	1 (0)	—	—	—
Division of Oral Maxillofacial Surgery	1 (0)	—	—	—

[a]AI: artificial intelligence.

Discussion

Principal Results

For clinical coders, it is clear from the MDCs that AI-assisted coding can serve as a reference for disease systems. However, hospital administrators may require detailed statistical results from clinical departments to make judgments. In the individual clinical department analysis based on the Wilcoxon signed rank test, the Division of General Internal Medicine, the Department of Neurology, and the Division of Neurosurgery had the highest number of cases studied, but the statistical results were inconsistent with coder-assigned coding. However, in the κ coefficient test, the statistical results for the nervous system MDC were highly consistent. This is because patients admitted to the Department of Neurology and Neurosurgery do not exclusively have neurological conditions. According to further analysis shown in Table 4, the AI model's predictions for neurological system diseases were still highly consistent with those of the disease classification staff in the neurology department. However, respiratory and urinary system conditions have affected the AI model's coding performance for neurology and account for the discrepancies seen in both the Wilcoxon signed rank test and the κ coefficient test.

Table . Analysis of major diagnostic category for 366 neurology patients admitted from February to April 2023.

No.	MDC[a]	AI[b]-assisted case coding (n=366), n (%)	Cases coded by human coders (n=366), n (%)	Kappa value	*P* value
1	Diseases and disorders of the nervous system	281 (76.8)	289 (79)	0.62	<.001
2	Diseases and disorders of the eye	17 (4.6)	18 (4.9)	0.84	<.001
3	Diseases and disorders of the ear, nose, mouth, and throat	21 (5.7)	20 (5.5)	0.77	<.001
4	Diseases and disorders of the respiratory system	1 (0.3)	1 (0.3)	–0.03	.96
5	Diseases and disorders of the circulatory system	20 (5.5)	12 (3.3)	0.54	<.001
6	Diseases and disorders of the digestive system	1 (0.3)	1 (0.3)	1.00	<.001
8	Diseases and disorders of the musculoskeletal system, and connective tissue	6 (1.6)	8 (2.2)	0.56	<.001
9	Diseases and disorders of the skin, subcutaneous tissue, and breast	2 (0.6)	2 (0.6)	1.00	<.001
10	Diseases and disorders of the endocrine, nutritional, and metabolic systems	1 (0.3)	2 (0.6)	0.67	<.001
11	Diseases and disorders of the kidney and urinary tract	4 (1.1)	1 (0.3)	0.40	<.001
14	Pregnancy, childbirth, and puerperium	0 (0)	1 (0.3)	—[c]	—
16	Diseases and disorders of the blood and blood forming organs and immunological disorders	1 (0.3)	1 (0.3)	1.00	<.001
17	Myeloproliferative diseases and disorders (poorly differentiated neoplasms)	0 (0)	1 (0.3)	—	—
18	Infectious and parasitic diseases and disorders	4 (1.1)	0 (0)	—	—
19	Mental diseases and disorders	3 (0.8)	0 (0)	—	—
23	Factors influencing health status and other contacts with health services	1 (0.3)	9 (2.5)	–0.05	.87
—	None	3 (0.8)	0 (0)	—	—

[a]MDC: major diagnostic category.

[b]AI: artificial intelligence.

[c]Analysis was not performed when there were 0 cases in a coding group.

In the circulatory system, the statistical results for the Division of Cardiology and Division of Cardiovascular Surgery in the Wilcoxon signed rank test were also markedly different. Upon closer examination of the data from the exploratory study, it was discovered that in the Division of Cardiovascular Surgery, half of the cases helped by AI-coding modules did not have the main diagnosis coded, which could be attributed to differences in how physicians document medical records. For example, after carefully reviewing the 24 cases of data collected by the Cardiac Surgery Department, it was found that 12 Taiwan DRGs were inconsistent. All of these did not follow the disease classification coding rules and did not include the main diagnosis (Table 5).

Table . Discussion on writing medical records in the Division of Cardiovascular Surgery.

Number	Excerpt of discharge diagnosis	Cause analysis
Case 1	"Chest tightness for a week Acute heart failure with reduced ejection fraction * 2023/04/24 Thallium 201 (Stress SPECT (single-photon emission computed tomography) imaging): mild myocardial ischemia in inferolateral wall of LV (left ventricle)"	The examination results were attached to the discharge diagnosis, leading to coding confusion.
Case 2	"Type A aortic dissection - post TEVAR (Thoracic endovascular aneurysm repair)+ stenting grafts in the ascending to descending aorta, left common carotid and subcalvian arteries on 2023/02/07 # Suspect gastroparesis related to relative gastric malperfusion"	The discharge diagnosis showed coding confusion with previously treated conditions.
Case 3	"Type B dissection, intramural hemorrhage - 2023/03/10 Chest CT (computed tomography) angiography:1) Suspect intramural hematoma in the descending aorta.2) Suspect thrombus formation in the bilateral femoral arteries.3) Suspect a thrombosed aneurysm in the right internal iliac artery"	The anatomical location was not clearly documented, making correct coding impossible.

Furthermore, in MDC 14 (pregnancy, childbirth, and puerperium) and MDC 21 (injuries, poisonings, and toxic effects of drugs), there are specific coding rules. Clinical coders need to synthesize the entire medical record information and apply the coding rules, which could result in diagnoses different from those presented in the discharge summary.

Limitations

The AI-coding module was trained on inpatient data from April 2019 to December 2020. Advancements in medical care might lead to variations in the diseases of admitted patients. Taken together, these show situations where the AI-coding assistance module might not capture the main diagnosis, as observed in the Dermatology Department.

Conclusions

With the rapid advancements in global medical technology and the evolving challenges of diseases, the development of DRG-based hospital payment systems in various countries is also meeting significant challenges. Key areas for future research include determining the flexibility of DRG payments, balancing payment structures, and aligning with disease management goals [23]. The Taiwan DRGs system, like those in other countries, aims to prevent medical institutions from delivering excessive services and causing unnecessary waste, all while safeguarding patient rights. It looks to strengthen management mechanisms to improve the quality and efficiency of care and ensure fair payments among peers.

In this context, AI-assisted coding emerges as a powerful tool [24]. A recent study used cross-random control methods to prove that AI-assisted coding reduces the coding workload [25]. The focuses of our research were the practical applications of AI models, with two main goals. The first was to investigate the consistency between the AI-assisted coding module and coding professionals and the second was to find the departments suitable for using the AI-assisted coding module. The research results showed that the highest consistency in MDC classification was seen in diseases of the respiratory system, as well as infectious and parasitic diseases. In the analysis of various inpatient specialties, departments such as the Division of Cardiology, Division of Nephrology, and Department of Urology showed no significant difference from coder-assigned coding results; accordingly, consideration could be given to integrating the AI-assisted coding module into the hospital information system, allowing physicians to reference Taiwan DRGs assignments for hospitalized patients, thus effectively controlling medical expenses.

However, upon analyzing the entire hospital department, discrepancies were observed in alignment with disease categorizations and personnel coding, so the research team is actively working on continuous improvements. Nevertheless, AI-assisted coding indeed served as a valuable reference by reducing human errors, as during the research period, it was found that the error rate detected by human coders (number of coding errors by human coders/total cases) was 1.9% (50/2632). Given the regular updates to the tool book by the Department of Health and the revisions in coding rules, the coding assistance module undoubtedly proves to be a powerful tool.

The development of AI-assisted coding for the *ICD-10-CM* PCS is just the beginning for intelligent health care in disease classification. Many operational aspects of hospitals are closely related to the *ICD-10-CM* PCS, including inpatient coding monitoring, discharge preparation services, and infectious disease surveillance, among others. For hospital administrators, the goal of AI-assisted coding is to achieve best operational revenue. For human coders with the assistance of an *ICD-10-CM* AI coding system, work time is reduced. Additionally, strengthening knowledge in clinical documentation improvement enables human coders to maximize their role, positioning them to become documentation experts [15] and preparing them for further career development.

Acknowledgments

This study was supported by grants from Kaohsiung Medical University Hospital (SI11315).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distribution of *ICD-10* (*International Statistical Classification of Diseases, Tenth Revision*) codes.
[JPEG File, 41 KB - humanfactors_v12i1e59961_app1.jpeg]

References

1. Hirsch JA, Nicola G, McGinty G, et al. ICD-10: history and context. AJNR Am J Neuroradiol 2016 Apr;37(4):596-599. [doi: 10.3174/ajnr.A4696] [Medline: 26822730]
2. Ayub S, Scali ST, Richter J, et al. Financial implications of coding inaccuracies in patients undergoing elective endovascular abdominal aortic aneurysm repair. J Vasc Surg 2019 Jan;69(1):210-218. [doi: 10.1016/j.jvs.2018.04.027] [Medline: 29937283]
3. Toner E, Khaled A, Ramesh A, Qureshi MK, Al Suyyagh K, Dunkow P. Financial impact of inaccurate coding plus cost-effectiveness analysis for surgically managed patients with periprosthetic fractures. Cureus 2021 Feb 1;13(2):e13060. [doi: 10.7759/cureus.13060] [Medline: 33680602]
4. Otero Varela L, Doktorchik C, Wiebe N, et al. International Classification of Diseases clinical coding training: an international survey. Hlth Inf Manag 2024 May;53(2):68-75. [doi: 10.1177/18333583221106509] [Medline: 35838185]
5. U.S. medical coding market size, share & trends analysis report by classification system (ICD, HCPCS, CPT), by component, by medical specialty, and segment forecasts, 2023 - 2030. Grand View Research. URL: https://www.grandviewresearch.com/industry-analysis/us-medical-coding-market [accessed 2025-01-06]
6. Venkatesh KP, Raza MM, Kvedar JC. Automating the overburdened clinical coding system: challenges and next steps. NPJ Digit Med 2023 Feb 3;6(1):16. [doi: 10.1038/s41746-023-00768-0] [Medline: 36737496]
7. Tong S, Jia N, Luo X, Fang Z. The Janus face of artificial intelligence feedback: deployment versus disclosure effects on employee performance. Strat Manag J 2021 Sep;42(9):1600-1631. [doi: 10.1002/smj.3322]
8. Vrontis D, Christofi M, Pereira V, Tarba S, Makrides A, Trichina E. Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. Int J Hum Resour Manag 2022 Mar 26;33(6):1237-1266. [doi: 10.1080/09585192.2020.1871398]
9. Kister TS, Remmler J, Schmidt M, et al. Acute kidney injury and its progression in hospitalized patients-results from a retrospective multicentre cohort study with a digital decision support system. PLoS ONE 2021;16(7):e0254608. [doi: 10.1371/journal.pone.0254608] [Medline: 34252151]
10. Wu S, Roberts K, Datta S, et al. Deep learning in clinical natural language processing: a methodical review. J Am Med Inform Assoc 2020 Mar 1;27(3):457-470. [doi: 10.1093/jamia/ocz200] [Medline: 31794016]
11. Chen PF, Wang SM, Liao WC, et al. Automatic ICD-10 coding and training system: deep neural network based on supervised learning. JMIR Med Inform 2021 Aug 31;9(8):e23230. [doi: 10.2196/23230] [Medline: 34463639]
12. Chraibi A, Delerue D, Taillard J, Chaib Draa I, Beuscart R, Hansske A. A deep learning framework for automated ICD-10 coding. Stud Health Technol Inform 2021 May 27;281:347-351. [doi: 10.3233/SHTI210178] [Medline: 34042763]
13. Diao X, Huo Y, Zhao S, et al. Automated ICD coding for primary diagnosis via clinically interpretable machine learning. Int J Med Inform 2021 Sep;153:104543. [doi: 10.1016/j.ijmedinf.2021.104543] [Medline: 34391016]
14. Chen Y, Chen H, Lu X, Duan H, He S, An J. Automatic ICD-10 coding: deep semantic matching based on analogical reasoning. Heliyon 2023 Apr;9(4):e15570. [doi: 10.1016/j.heliyon.2023.e15570] [Medline: 37151662]
15. Campbell S, Giadresco K. Computer-assisted clinical coding: a narrative review of the literature on its benefits, limitations, implementation and impact on clinical coding professionals. Hlth Inf Manag 2020 Jan;49(1):5-18. [doi: 10.1177/1833358319851305] [Medline: 31159578]

16. Asadi F, Ahmadi Hosseini M, Gomar T, Sabahi A. Factors affecting clinical coding errors. Shiraz E-Med J 2022;23(9). [doi: 10.5812/semj-122161]
17. Islam MM, Li GH, Poly TN, Li YCJ. DeepDRG: performance of artificial intelligence model for real-time prediction of diagnosis-related groups. Healthcare (Basel) -> Healthc (Basel) 2021 Nov 25;9(12):1632. [doi: 10.3390/healthcare9121632] [Medline: 34946357]
18. Shuai Z, Xiaolin D, Jing Y, et al. Comparison of different feature extraction methods for applicable automated ICD coding. BMC Med Inform Decis Mak 2022 Jan 12;22(1):11. [doi: 10.1186/s12911-022-01753-5] [Medline: 35022039]
19. Subotin M, Davis AR. A method for modeling co-occurrence propensity of clinical codes with application to ICD-10-PCS auto-coding. J Am Med Inform Assoc 2016 Sep;23(5):866-871. [doi: 10.1093/jamia/ocv201] [Medline: 26911826]
20. Chang NW, Dai HJ, Jonnagaddala J, Chen CW, Tsai RTH, Hsu WL. A context-aware approach for progression tracking of medical concepts in electronic medical records. J Biomed Inform 2015 Dec;58 Suppl(Suppl):S150-S157. [doi: 10.1016/j.jbi.2015.09.013] [Medline: 26432355]
21. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. Presented at: 2019 Conference of the North American Chapter of the Association for Computational Linguistics; Jun 2-7, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.18653/v1/N19-1423]
22. Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E. Hierarchical attention networks for document classification. Presented at: 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies; Jun 12-17, 2016; San Diego, CA p. 1480-1489. [doi: 10.18653/v1/N16-1174]
23. Liu Y, Wang G, Qin TG, Kobayashi S, Karako T, Song P. Comparison of diagnosis-related groups (DRG)-based hospital payment system design and implementation strategies in different countries: the case of ischemic stroke. Biosci Trends 2024 Mar 19;18(1):1-10. [doi: 10.5582/bst.2023.01027] [Medline: 38403739]
24. Wang H, Gao C, Dantona C, Hull B, Sun J. DRG-LLaMA: tuning LLaMA model to predict diagnosis-related group for hospitalized patients. NPJ Digit Med 2024 Jan 22;7(1):16. [doi: 10.1038/s41746-023-00989-3] [Medline: 38253711]
25. Chomutare T, Lamproudis A, Budrionis A, et al. Improving quality of ICD-10 (International Statistical Classification of Diseases, Tenth Revision) coding using AI: protocol for a crossover randomized controlled trial. JMIR Res Protoc 2024 Mar 12;13:e54593. [doi: 10.2196/54593] [Medline: 38470476]

Abbreviations

AI: artificial intelligence
BERT: bidirectional encoder representations from transformers
DRG: diagnosis-related group
HAN: hierarchical attention network
***ICD*:** *International Classification of Diseases*
***ICD-10*:** *International Statistical Classification of Diseases, Tenth Revision*
***ICD-10-CM*:** *International Classification of Diseases, Tenth Revision, Clinical Modification*
***ICD-9-CM*:** *International Classification of Diseases, Ninth Revision, Clinical Modification*
MDC: major diagnostic category
NHIA: National Health Insurance Administration
PCS: procedure coding system
WHO: World Health Organization

Edited by A Kushniruk; submitted 30.04.24; peer-reviewed by S Mao, YT Chang; revised version received 28.09.24; accepted 01.12.24; published 12.02.25.

Please cite as:
Lu AT, Liou CS, Lai CH, Shian BT, Li MT, Sun CY, Kao HY, Dai HJ, Tsai MJ
Application of Clinical Department–Specific AI-Assisted Coding Using Taiwan Diagnosis-Related Groups: Retrospective Validation Study
JMIR Hum Factors 2025;12:e59961
URL: https://humanfactors.jmir.org/2025/1/e59961
doi:10.2196/59961

Original Paper

Enhancing Early Language Disorder Detection in Preschools: Evaluation and Future Directions for the Gades Platform

María Dolón-Poza[1*], MSc; Ana-Marta Gabaldón-Pérez[1*], PhD; Santiago Berrezueta-Guzman[2*], PhD; David López Gracia[1*], MSc; María-Luisa Martín-Ruiz[1*], PhD; Iván Pau De La Cruz[1*], PhD

[1]Grupo de Investigación Innovación Tecnológica para las Personas (InnoTep), Departamento de Ingeniería Telemática y Electrónica, ETSIS Telecomunicación, Campus Sur, Universidad Politécnica de Madrid, Madrid, Spain

[2]Technical University of Munich, Heilbronn, Germany

[*]all authors contributed equally

Corresponding Author:
Santiago Berrezueta-Guzman, PhD
Technical University of Munich
BildungsCampus 2
Heilbronn, 74076
Germany
Phone: 49 15159184710
Email: s.berrezueta@tum.de

Abstract

Background: Language acquisition is a critical developmental milestone, with notable variability during the first 4 years of life. Developmental language disorder (DLD) often overlaps with other neurodevelopmental disorders or simple language delay (SLD), making early detection challenging, especially for primary caregivers.

Objective: We aimed to evaluate the effectiveness of the Gades platform, an adaptive screening tool that enables preschool teachers to identify potential language disorders without direct support from nursery school language therapists (NSLTs).

Methods: The study took place in a nursery school and an early childhood educational and psychopedagogical center in Madrid, Spain, involving 218 children aged 6 to 36 months, 24 preschool teachers, and 2 NSLTs. Initially, NSLTs conducted informational sessions to familiarize teachers with DLDs and how to identify them. Following this, the teachers used the Gades platform to conduct language screenings independently, without ongoing support from NSLTs. The Gades platform was enhanced to collect detailed profiles of each child and implemented an adaptive screening model tailored to account for variability in language development. This setup allowed preschool teachers, who are not language experts, to observe and assess language development effectively in natural, unsupervised educational environments. The study assessed the platform's utility in guiding teachers through these observations and its effectiveness in such settings.

Results: Gades identified language difficulties in 19.7% (43/218) of the children, with a higher prevalence in boys (29/218, 13.3%) than in girls (14/218, 6.4%). These challenges were most frequently observed in children aged 15 to 27 months. The platform demonstrated a high accuracy rate of 97.41%, with evaluators largely agreeing with its recommendations. Teachers also found Gades to be user friendly and a valuable tool for supporting language development observations in everyday educational settings.

Conclusions: Gades demonstrates potential as a reliable and accessible tool for early detection of language disorders, empowering educators to identify DLD and SLD in the absence of NSLTs. However, further refinement of the platform is required to effectively differentiate between DLD and SLD. By integrating Gades into routine preschool assessments, educators can facilitate timely interventions, bridging gaps in early childhood education and therapy.

Trial Registration: Pan-African Clinical Trial Registry (PACTR) PACTR202210657553944; https://pactr.samrc.ac.za/TrialDisplay.aspx?TrialID=24051

(JMIR Hum Factors 2025;12:e60424) doi:10.2196/60424

KEYWORDS

developmental language disorder; simple language delay; adaptive screening system; early childhood education; pervasive therapy

Introduction

Screening and Prevalence of Developmental Language Disorder

Communication plays a fundamental role in children's cognitive development. Through communication, children acquire knowledge, express their thoughts and emotions, develop their cognitive ability, and establish relationships with others. Adequate communication development enables them to learn and participate in various social contexts. Therefore, language acquisition during childhood is one of the most critical developmental milestones but with the most substantial interindividual variability in the first 4 years of life. Children start consolidating their language learning process from the age of 4 to 6 years [1-7].

The *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*), published by the American Psychiatric Association, introduced a new category known as neurodevelopmental disorders (NDDs) that groups different conditions that emerge early in child development before starting elementary school. These conditions eventually persist throughout adulthood [8-12].

The NDDs include communication disorders related to language, speech, and communication impairments. Developmental language disorder (DLD) is classified as a communication disorder. Children with DLD experience challenges in the acquisition and use of language. Language proficiency depends on receptive and expressive abilities. Therefore, children diagnosed with DLD encounter difficulties in both domains [12]. Thus, vocabulary is more limited than expected, and sentences are shorter and less complex with grammatical errors or with speech alterations in narration and comprehension or production of sentences. In addition, they may appear shy or prefer to communicate only with their family members. Children with DLD not only have symptomatology in communication skills but also have difficulties in cognitive and sociocultural processes [5,13-16].

Moreover, a family history of language disorders is often present, and DLD occurs with other disorders, such as autism spectrum disorder and attention-deficit/hyperactivity disorder (ADHD), in 30% of cases [5,17,18]. This makes early detection of DLD in these children more challenging for primary care pediatricians due to the significant comorbidity and similar symptoms with other NDDs [19].

On the other hand, early detection of DLD can be masked by misidentifying it as simple language delay (SLD). DLD is a persistent disorder with a slow rate of improvement and significant variability. This differentiates it from SLD [1,5,20]. However, this indicator does not allow for early detection of DLD, as we cannot identify whether the child has a one-off delay or is showing symptoms of DLD. To differentiate among them, difficulties must be identified in other areas, such as semantics, pragmatics, morphosyntax, and phonology [3]. Therefore, language difficulties manifest through the aforementioned skills that are measurably below what is expected for age, significantly impacting academic achievement, work performance, communication, and socialization [8-11].

DLD is diagnosed at the age of 4 years and is usually stable over time, as interindividual differences in language ability are reduced [1]. However, there may be warning signs that manifest themselves earlier. A study carried out by Vall d'Hebron Hospital Universitari in Barcelona, Spain, highlights a significant portion of children aged 5 to 17 years who exhibit clear symptoms of NDDs but have not been previously diagnosed with DLD [21]. Moreover, communication disorders are one of the significant NDDs with more prevalence in Spanish schools (1.05%-3.42%), along with ADHD and learning disorders [21,22]. Other studies estimated that the population prevalence of language disorders, without any relation to other intellectual disabilities, is 7.58% for children aged 4 to 5 years [17] and 6.4% at the age of 10 years [23]. To summarize, within a preschool setting of 30 children, it can be anticipated that about 2 children may exhibit DLD manifestations [17]. In total, 10% of preschool children present difficulties in language acquisition, of whom 5% to 7% end up being diagnosed with DLD, and the rest are diagnosed with SLD [5].

Before the detection and diagnosis of DLD, there is a prevention phase. Here, the nursery school plays a key role, as it is within this environment that the child spends most of their time and is involved in different interactions [1]. These interactions contribute to developing their communicative skills (communicative intention, nonverbal communication, imitation, waiting, etc) with peers of a similar age and teachers. Research has demonstrated a correlation between children with DLD and their academic performance. In fact, 88% of these children fail to meet the necessary curricular standards during their first year of school [14,17]. Hence, there is a pressing need to support education professionals with enough training and knowledge to identify early indicators of potential language disorders [1,24-27].

In the detection phase, the aim is to identify early whether a child is suspected of having some NDDs. To achieve adequate detection, tools such as screening forms and questionnaires allow information to be compared with previously defined risk indicators [20,28-32]. Some regions, such as Madrid (Spain), have early childhood educational and psychopedagogical guidance teams supporting nursery schools' detection phase. These teams engage in preventive actions and collaborate to detect developmental issues during the initial years of a child's life. If an educational case is identified, it is referred to the early intervention centers responsible for the assessment and intervention phase [33].

The assessment and intervention cover different aspects. During the assessment phase, it is necessary to analyze the family environment through interviews with relatives. In addition, hearing tests should be administered to eliminate potential associated psychological disorders. Evaluation of communication, playing with peers, comprehension, language production, and psychomotor skills are also essential. Finally, this information is correlated with medical history [6,14,15,20]. The intervention phase for DLD should be multidisciplinary, involving teachers, occupational therapists, speech therapists,

and psychologists skilled in children's language development techniques [7,24,27,34].

Gades: A Screening Tool for DLD

Gades functions as a clinical decision support system designed to grant expert knowledge regarding language disorders to preschool teachers. It offers a screening form to assess the language acquisition process during the early stages of a child's development. Gades allows the implementation of screening systems in natural environments for the child, thereby embracing the principles of pervasive therapy [35].

Pervasive or ubiquitous therapy is a therapeutic approach supervised by professionals that aims to promote engagement in activities focused on improving the individual's life within their natural environment through the support of information and communication technologies. It aims to be person centered instead of focusing on a specific clinical pathology. It especially seeks to identify how a problem impacts and manifests across various contexts [18,36]. This approach has the same phases of prevention, detection, assessment, and intervention as normative therapy. Moreover, using information and communication technologies ensures consistency and continuity of the therapeutic process within the person's natural setting [32,37-39].

For children's pervasive therapy, its application extends beyond the clinical center to environments like the child's school or home. Nursery schools are included in these settings. Being places where children spend much of their time, they are crucial for the early detection of DLD [24]. During this educational phase, teachers or educators and specialists in therapeutic pedagogy, such as nursery school language therapists (NSLTs), are typically involved. However, many nursery schools lack NSLTs, and even when they are available, fewer than half of the children suspected of having a language disorder are referred to or receive support from these specialists [1,7,17,23,27].

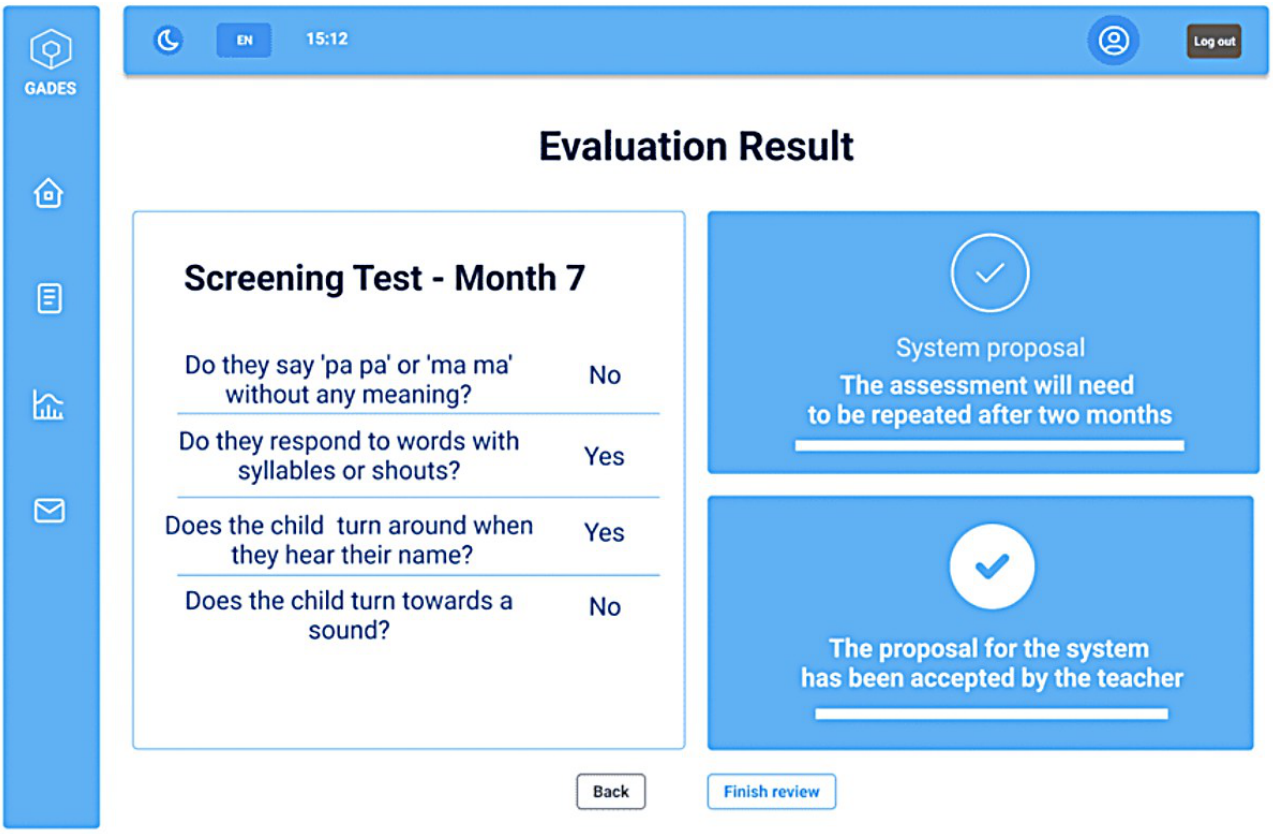
Martín Ruiz et al [35] developed and evaluated a previous version of Gades. The platform's cornerstone is a knowledge base (KB) that includes 106 milestones related to children's language acquisition based on age, defined and agreed upon by language experts. The previous version of this KB covered children aged 1 month to 72 months [35]. Building upon the findings of the previous experiment, increased instances of suspected DLD were identified in children aged 0 to 3 years. Consequently, in the current experiment, we focus on prioritizing the implementation of Gades during this developmental period. To extend this goal, Gades has undergone modifications and enhancements to its functionalities to gather additional pertinent information concerning language acquisition in children. These adaptations include integrating inquiries specific to the bilingual context of the children and their family backgrounds and evaluating results from the Gades across different language areas.

When using Gades, the corresponding milestones, tailored to the child's condition, are incorporated into a screening form through questions that education professionals use to assess a child's language development level. The milestones are categorized into 2 main types: "warning milestones," which prompt a reevaluation, and "alarm milestones," indicating the need for direct referral to health professionals. Gades' KB assesses 4 areas of speech and language development: sensory reception, speech perception, production, and pragmatic.

After the form adapted to the child is completed, Gades generates an evaluation result, including a suggestion for action if deviations from typical child development stages are detected. Therefore, it provides education professionals with expert knowledge on language acquisition, enhancing the early detection of DLD. Furthermore, this screening form acts as a psychopedagogical report for early intervention health professionals. Gades can aid in identifying the language development profile of a child suspected of having DLD through the assessment of processing-oriented and performance-based tasks.

Figure 1 depicts the graphical user interface of Gades, showcasing the evaluation results of consultation for a child aged 7 months. The platform displays the outcomes of the screening form, including the posed questions and the teacher's responses. Response options include "yes" if the child performs the specified action, "no" if not, and "do not know/no answer" for unknown answers. This form is completed based on the preschool teacher's observations of the child without direct interaction. After review, the system recommends that the assessment be repeated in 2 months, a suggestion that the overseeing teacher has validated. Gades also features additional tools for child registration, language evaluation, and user profile management.

Figure 1. Gades' evaluation result page for a child aged 7 months. In total, 4 questions were asked based on the child's age, and the system suggested repeating the evaluation in 2 months. The preschool teacher concurred with the system's recommendation.



Goal of This Study

For effective early detection and intervention of DLD, it is essential to offer sufficient services to these children to minimize the impact on their social, emotional, and educational development. Thus, language disorders often require a multidisciplinary approach among professionals from diverse fields such as education, speech-language therapy, and medicine to extend their reach into various settings [17,18,24]. Supporting training programs and systems for health and education professionals fosters collaboration. This collaboration facilitates the creation of adequate evaluation and intervention plans that meet the diverse needs of individuals with language difficulties, enhancing the quality of care and inclusivity [22,26,30,31,38,40].

Within this collaboration, the early detection of warning signs, which are potential indicators of language disorders in children, should be an integral part of the daily work of those interacting with children [1,7]. Supporting early childhood education professionals is crucial to ensure that they possess adequate knowledge and tools for preventing and detecting early language development issues in children aged 0 to 3 years because behaviors and warning signs related to language disorders can be identified from the age of 2 years onward [1,3,4,24]. It facilitates appropriate early diagnosis of DLD or SLD through subsequent assessment at early intervention centers. Moreover, given the absence of an NSLT in some schools, preschool teachers must have access to these resources.

Considering the higher incidence of suspected DLD observed in children aged 0 to 36 months in the previous experiment [17,23], this paper aims to emphasize the use of Gades during this period, targeting kindergarteners aged 6 to 36 months in a nursery school and an early childhood educational and psychopedagogical center. Gades has been adapted and expanded to gather more relevant information on language acquisition, including questions tailored to bilingual contexts and family backgrounds. Therefore, our primary goal is to evaluate Gades' ability to differentiate between typical and atypical language development for the target ages with these minor adjustments.

As the previous experiment consisted of a controlled environment, this paper also presents the relationship between preschool teachers and the Gades platform. Specifically, how Gades serves as a valuable tool for guiding nonexperts, such as preschool teachers, in conducting language observations. Therefore, it evaluates Gades' effectiveness in a real setting and its adherence to clinical criteria without direct supervision or intervention of language experts or technical staff. A potential challenge is also associated with integrating these technological tools into unsupervised educational settings to support therapeutic interventions [41].

Moreover, this screening process aims to identify the patterns indicative of a child with problems in language development, including SLD or DLD. Considering the updated *DSM-5* criteria, language development in children occurs within heterogeneous and dynamic contexts. Hence, it is imperative to use screening tools to identify significant patterns and milestones in the language development of a child with DLD, facilitating early detection and intervention. To accomplish this, the screening

tools must be intersectional, considering influential parameters, such as clinical conditions and the sociocultural and linguistic environment surrounding the child. Consequently, the final goal is to outline the framework for an adaptive screening model for language disorders that reflects the child's current developmental and social context. This approach ensures a comprehensive and nuanced understanding of each child's unique situation, enhancing the accuracy and effectiveness of early detection methods. Therefore, this study hypothesizes that the Gades platform can effectively support preschool teachers in the early detection of language development difficulties, including DLD and SLD, within natural educational settings, achieving high accuracy rates and providing actionable insights without requiring direct supervision from language disorder specialists.

Methods

Overview

This study used a prospective observational design to evaluate the updated Gades platform in real-world educational settings. The design involved monitoring children aged 6 to 36 months and their evaluators—preschool teachers and early childhood educators—over a defined period, collecting data as they naturally occurred during regular classroom activities. This approach was selected to ensure that the platform's performance could be assessed in authentic, uncontrolled conditions, reflecting its practical application. By focusing on how educators integrated Gades into their daily routines, the study aimed to capture unbiased insights into its usability and effectiveness in identifying early signs of language development difficulties.

Requirement Analysis

Individuals encounter varied communicative situations involving diverse entities and methods in everyday life. Current diagnoses and treatments for various disorders frequently overlook key characteristics, such as diversity, heterogeneity, and the role of socialization. The *DSM-5* and numerous studies highlight the necessity to consider interindividual differences in language acquisition early on. These sources suggest that assessments of language, speech, and communication should consider sociocultural aspects, linguistic context (including dialects), and socioeconomic status, as these aspects directly influence language development [1,12,16,17,19,21,31,32,42-44].

Bosch et al [21] demonstrated that 1.05% of the participants had communication disorders, with the most significant factors being foreign origin, genre, socioeconomic status, and age. Consequently, the diagnosis of language impairment should consider the individual's background, bilingual context, direct clinical observations in settings such as home or school, and results of standardized tests to assess the severity of the disorder.

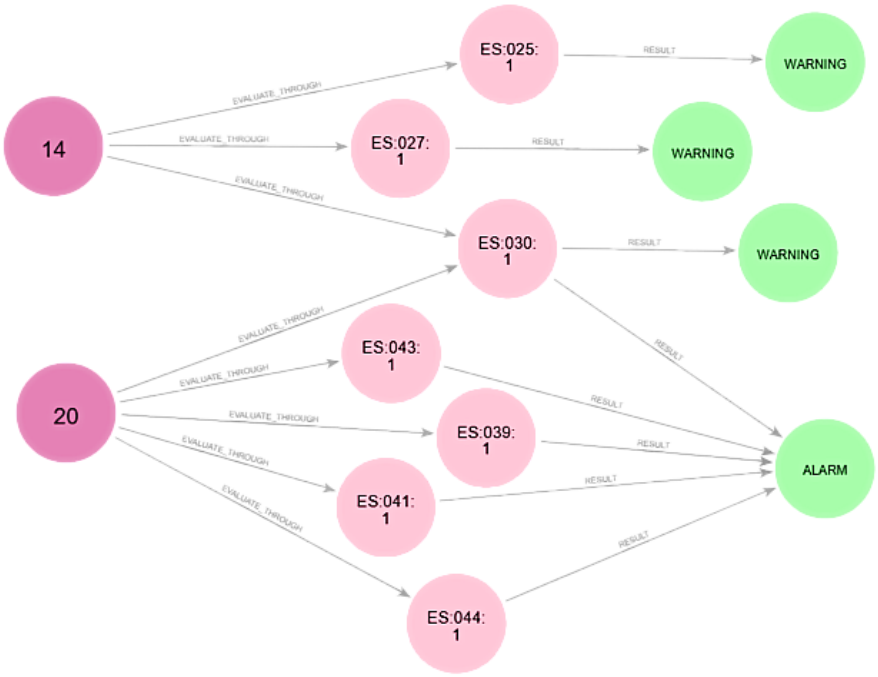
The New Gades Platform

Building on the previous experiment with the Gades platform [35], we updated it with new technologies and features. These enhancements enable the collection of additional information from children to conduct more accurate and suitable assessments of a child's language development. This includes gathering details about the child's family background, linguistic environment, and medical history, all of which contribute to a more comprehensive understanding of each child's unique developmental context. Therefore, incorporating these minor adjustments delineates the structure of an adaptable screening model that considers the variability in language development. This methodology guarantees a thorough and nuanced comprehension of each child's individual circumstances, enhancing the precision and efficacy of techniques for identifying potential issues in language development.

Implementing Neo4j (Neo4j Inc), a graph database technology, is crucial in optimizing our Gades platform, specifically in enhancing the accuracy of language disorder detection. Neo4j allows for a more natural and flexible representation of complex data and interconnected relationships between them, which is essential for modeling the intricate interactions and dependencies observed in child language development [45]. In the context of Gades, Neo4j facilitates the dynamic integration and analysis of large volumes of data related to language development milestones, assessment responses, and individual characteristics of children. For example, the graph structure enables us to directly link children's responses to specific questions with language development patterns and automatically flag connections that indicate potential delays or deviations from typical development.

Therefore, the KB has also been migrated to Neo4j graph databases. As illustrated in Figure 2, each milestone is represented through a month (dark pink node) with several questions (pale pink nodes) and a final suggestion (green nodes). Thus, in the figure, a dark pink node is represented each month, corresponding to different questions designed to evaluate the child's language proficiency, depicted by pale pink nodes. Notably, some questions are linked to multiple months, reflecting the dynamic nature of language development assessment. Different action suggestions can be generated based on the answers to these questions, such as confirming typical development, recommending reassessment, or advising referral to early childhood care services (green nodes represent these). This graph-based approach facilitates a nuanced analysis of the child's language development, allowing for tailored interventions sensitive to each child's unique progression.

Figure 2. Gades' knowledge base in Neo4j: milestone months (dark pink node), questions identifiers (pale pink nodes), and suggestions (green nodes). Each month has various questions to evaluate language acquisition (for a child aged 14 months or 20 months) with suggestions such as typical development, repeat evaluation, or advising referral to early childhood care services.



With the updates to Gades, the system has addressed feedback from educators seeking more information on DLD. Previously, questions were categorized by language development areas and were only accessible to the system's technical staff. However, with the recent improvements, information about various skills is now accessible to the evaluation staff. This improvement means that by using Gades, preschool teachers and other professionals can gain deeper insights into the specific language development areas being assessed. Such transparency not only empowers educators with a better understanding of language development but also enables them to contribute more effectively to detecting warning signs.

Table 1 illustrates the classification and number of questions across different categories. The sensory reception category is administered in the earliest months of life, primarily to assess for hearing issues, and gradually transitions into language perception. This aspect entails the capacity to receive, process, and comprehend linguistic information through sensory channels, whether auditory or visual. Pragmatics includes conversational skills, coherence, cohesion within discourse, and the functional and social use of the language. Finally, language production encompasses phonology, morphosyntax, semantics, and nonverbal communication. This structured approach allows for a comprehensive assessment of a child's language development, targeting specific areas crucial for early detection and intervention of language disorders.

Table 1. Classification of questions according to language category [35].

Language development domains	Number of questions
Sensory reception	3
Language perception	32
Production	48
Pragmatics	25

Hypothesis and Experiment

We aimed to validate the reliability and effectiveness of the new Gades platform in a real setting without the direct supervision or intervention of language experts or technical staff. We explored the interaction between preschool teachers and the Gades platform, highlighting how Gades functions as a valuable tool for facilitating language observations to detect warning signs of language difficulties by nonexperts. The Gades platform is specifically designed to assist in detecting early language difficulties that may signal potential DLD. By focusing on these early warning signs, Gades supports the early detection and monitoring of language development challenges in real-world educational settings.

The Gades KB was established in a previous study [35] and is grounded in standardized tests commonly used to assess DLD. This ensures that Gades relies on validated methodologies to identify and interpret potential language difficulties.

The study's population sample consisted of 218 children aged 6 to 36 months from 2 educational centers in the city of Madrid: a preschool center and an early childhood educational and

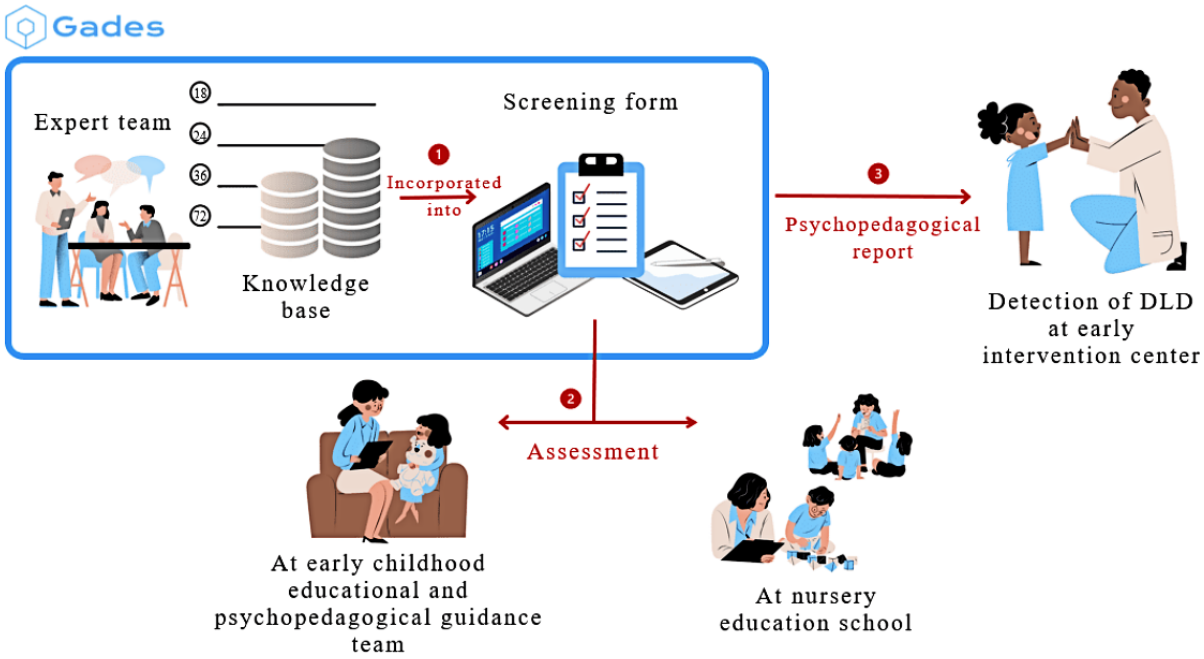
psychopedagogical center. All children enrolled in the first cycle of early childhood education, which aligns with this age range, participated in the study. This selection was based on recent evidence highlighting the critical importance of identifying language difficulties during this developmental period [1,3,4,24]. The screening data were collected from March to June 2023, providing an initial assessment of Gades' effectiveness in a real-world educational setting. Notably, access was limited to Spanish-speaking and some bilingual children.

The screening assessments for DLD involved the active participation of 24 nursery school educators and preschool teachers and a clinical professional from the early childhood center. Throughout this study, we will refer to them as evaluators. Hence, 2 categories of evaluators were involved in the experiment. The initial cohort consisted of professionals from early childhood centers, who, while not specialized in DLD, possessed adequate knowledge of NDDs. The second group comprised nursery school educators and preschool teachers, who may lack specialized language expertise but possess valuable insights into the children's competencies under their care. Before starting the experiment, the evaluators received training from qualified professionals affiliated with the Specific Language Disorder Association of Madrid. This training aimed to deepen their understanding and expertise regarding DLD and its identification with children aged 6 to 36 months. The training, delivered in a concise 2-hour format, was designed to ensure efficiency while providing comprehensive insights. By the session's conclusion, it was anticipated that the educators would have developed a more nuanced understanding of typical language milestones and a heightened capacity to discern early indications of language difficulties in children under their care.

During the evaluations, they operated without direct assistance from the training professionals or the Gades technical support team. Consequently, the experiment was conducted in an uncontrolled setting, relying exclusively on the evaluators' direct observation of the children's behavior for further evaluation through Gades. The evaluations were conducted following the guidelines recommended by Gades, aiming to ensure that the process was as effective and informative as possible within the constraints of the study's design.

Figure 3 outlines the Gades' framework, starting with an initial phase where a series of meetings were conducted to review and validate the KB established in the previous experiment. This phase involved experts in language disorders and the technical personnel from Gades to ensure the KB's relevance and accuracy. The panel of experts, who were also involved in constructing the KB as part of a previous study [35], recommended that including 3 to 6 questions per month would constitute an adequate measure for detecting language development issues. In the context of this study, the same experts conducted a review and confirmed the continued validity and applicability of the content of the KB. Once the KB was validated, the second phase, the screening process, began.

Figure 3. Gades' detailed framework. The process starts with expert validation of the knowledge base, integrated into a screening form with control and language questions. Evaluators assessed the children, and Gades provided recommendations, such as reassessment or referral, forming the basis of a psychopedagogical report. DLD: developmental language disorder.



The assessments were carried out by experienced female educators and professionals proficient in using digital platforms and information technologies. They used mobile devices and workplace computers to conduct evaluations, strategically scheduling them during the children's nap time around midday to minimize distractions. This approach involved each evaluator conducting assessments of the children under her care, which consisted of 2 parts. The first part included control questions related to the child's medical history and linguistic context, serving as a baseline for understanding each child's unique background. The second part consisted of the language difficulties screening questions. These questions were

meticulously developed in collaboration with experts to cover risk parameters for the early detection of DLD.

After the screening form was completed, the Gades system generated a recommendation to the evaluator, such as suggesting a repeat assessment in the upcoming months or indicating the need for a referral to an early intervention center. If a referral was recommended, nursery schools in Madrid forwarded the result to the expert through a psychopedagogical report. The designated expert, a language professional, was responsible for conducting a standardized evaluation tailored to the child's needs. Meanwhile, the educator played a supportive role by closely observing the child's progress in class and paying particular attention to areas of difficulty. Regular communication between schools and the expert was maintained with biweekly visits to ensure follow-up and support.

The application of the Gades platform was deliberately structured to be as nonintrusive as possible, allowing evaluators to focus on their primary duties while still contributing to the study. On average, each child underwent 1.24 (SD 0.49) assessments, indicating a consistent approach to the number of evaluations. The average duration of an assessment was 70 (SD 41.8) seconds per child, demonstrating the platform's efficiency in collecting data within a brief timeframe. To validate the Gades platform, both quantitative and qualitative analyses were conducted.

Quantitative validation focused on assessing the platform's accuracy and reliability. Descriptive statistics were used to summarize the duration and frequency of assessments. Agreement between the system's recommendations and evaluators' decisions was measured using concordance rates, calculated as the proportion of evaluations where the evaluators accepted Gades' recommendations. Statistical significance was set at $P<.05$. In addition, the platform's accuracy rate (97.41%) was determined by dividing the number of accepted suggestions (n=263) by the total number of evaluations (N=270). Instances of disagreement were analyzed qualitatively to identify patterns or recurring issues.

Qualitative validation focused on the integration of Gades into educational settings and its usability as perceived by preschool teachers. Evaluators were asked to provide feedback via a standardized questionnaire, following a validated system usability scale [41]. Responses were anonymized and collected outside the Gades platform to reduce potential bias. Evaluators also provided open-ended feedback to capture their experiences and insights, highlighting the platform's strengths and areas for improvement. The qualitative data were analyzed thematically to identify trends and recurring themes.

This combination of statistical and thematic analyses ensured a comprehensive validation of Gades. By comparing the platform's outcomes against educator feedback and established clinical criteria, the study aimed to validate Gades' ability to aid in the early detection of language development issues, such as SLD or DLD. Furthermore, comparisons between the initial version of Gades (2014) [35] and the current version highlighted improvements in adherence to clinical standards and system usability.

Ethical Considerations

This study, including the validation of the Gades platform and the experiment conducted through it, received ethics approval from the Ethics Committee of the Polytechnic University of Madrid in October 2023 GYGSDAALTD-MLMR-DATOS-20231010 and GYGSDAALTD-MLMR-HUMANOS-20231010). This approval ensures that both the use of the Gades platform and the specific experimental protocol comply with ethical standards, including data protection, participant rights, and research transparency.

For this study, an opt-out consent model was implemented. Participants were informed in advance about the study's objectives, the nature of the data collection, and the measures taken to ensure anonymity. They were given the opportunity to decline participation or withdraw their data at any point without providing a justification. Consent was implied by the participants' choice to continue with the evaluations, as no identifiable personal data were collected. Anonymized data were securely provided to the research team.

The professionals overseeing the evaluations maintained confidentiality in accordance with professional secrecy standards. Only the principal investigator had access to the evaluation results, which were securely stored in an encrypted database, accessible solely through authorized credentials.

Results

Sample Summary

The sample size for this study was calculated using a finite population correction formula to ensure adequate representation and statistical power. On the basis of demographic data from Madrid, the estimated population of children aged 6 to 36 months was approximately 150,000 [46]. With an assumed prevalence of DLD of 7%—a conservative estimate derived from epidemiological studies [17]—a 95% CI (Z=1.96), and a margin of error of 5%, the required sample size was calculated. The formula incorporates the finite population adjustment to account for the limited size of the target population, yielding a minimum required sample size of approximately 100 participants. This calculation also considered the variability in language development within the population.

The study's population sample comprised 218 Spanish-speaking children aged 6 to 36 months, more than double the calculated minimum sample size, ensuring robust statistical power and enhancing the reliability of the results. It featured a control group of 134 children from the nursery school (center 1) without any previous diagnosis or visible language difficulties and an experimental group of 84 children from the early childhood educational and psychopedagogical center (center 2), suspected of NDDs. Both groups were matched by gender to maintain homogeneity. Each child was assigned a random identifier for data anonymization and to facilitate subsequent analysis by the research team.

In Madrid's nursery schools, for each child, the evaluations were conducted by 2 assigned preschool teachers, also called an "educational pair." The concept of an "educational pair"

refers to 2 professionals sharing responsibilities and collaborating for the children's development and well-being, enabling the distribution of evaluation tasks between them. Therefore, in classrooms with children aged <12 months, each teacher evaluated an average of 4 (SD 0) children. For groups with children aged from 13 to 24 months, the responsibility increased to an average of 6.75 (SD 0.577) children per teacher. For those aged >24 months in the classroom, the average was 5.9 (SD 6.610) children per preschool teacher. Evaluations at the early childhood educational and psychopedagogical center were conducted by the teacher specializing in therapeutic pedagogy (NSLT).

Table 2 presents an overview of the study's sample, detailing the distribution of the children by age and gender: 45.4% (99/218) of the girls and 54.6% (119/218) of the boys participated. Within this group, 36% (36/99) of the girls and 40.3% (48/119) of the boys came from the early childhood center, initially identified with potential NDD concerns. Specifically, for participants aged >3 years, the study included 4 boys and 2 girls, all aged 37 months, except for 1 girl who was aged 38 months. The second year of the initial preschool cycle had the highest representation, with the largest number of children aged between 18 and 28 months, averaging 10 (SD 3.045) children per month. The average age in months per grade was distributed as follows: up to 12 months, the average age was 10.020 (SD 1.136) months; from 13 to 24 months, the average age was 20.331 (SD 1.108) months; and from 25 to 36 months, the average age was 31.578 (SD 1.094) months.

Table 2. Distribution of children by center, gender, and age range[a].

Grade	Up to 12 mo (n=26), n (%)	13-24 mo (n=102), n (%)	25-36 mo (n=84), n (%)
Children from center 1			
Female	10 (38.4)	22 (21.6)	31 (36.9)
Male	8 (30.7)	24 (23.5)	39 (46.4)
Children from center 2			
Female	5 (19.2)	22 (21.6)	9 (10.7)
Male	3 (11.5)	34 (33.3)	11 (13.1)

[a]Center 1 is the nursery school, and center 2 is a psychopedagogical center. The age ranges align with the preschool grades. The second year of the initial preschool cycle had the highest representation (102 children).

The screening form's initial section included questions regarding the child's medical history and linguistic context. Birth complication cases were more prevalent in center 1, particularly among male participants. Results indicated that 8.2% (11/134) of the children with no previous diagnosis of NDD experienced birth complications, including prenatal, perinatal, and preterm risks. 2.9% (4/134) were reported among females while among males, 5.2% (7/134) were observed. In center 2, only 2% (2/84) cases were reported in males and no cases in females. Notably, 3 children faced more than 2 of these risks, such as family history combined with perinatal risk. The mean gestational weeks remained relatively consistent across both centers and genders, with slightly more variation among girls. The average gestational age was 39.518 (SD 5.363) weeks, female children had an average of 38.984 (SD 2.36) weeks in center 1 and 40 (SD 0) weeks in center 2, while males had 39.563 (SD 1.29) weeks in center 1 and 39.791 weeks (SD 1.44) in center 2. A baby is considered premature if born before the 37th week of pregnancy. According to the study, only 3.7% (8/218) of the children were born preterm, highlighting specific early life factors that could influence developmental outcomes. Instances with a family history or bilingual cases were limited. Among the 218 cases studied, only 2 reported a family history of other NDDs such as autism spectrum disorder. Information regarding the bilingual context was exclusively available for children from center 1 (N=134). 7 (5.2%) cases were found among females, and 3 (2.2%) cases were recorded in males. It suggests that the language development impact of being raised in a bilingual environment could not be thoroughly assessed across the entire sample.

Overview of the Obtained Health Results

This section presents the results obtained from the evaluations that were conducted. A total of 270 assessments were completed, accounting for instances where children underwent multiple assessments. Retaining only the latest assessment conducted for each child (218 evaluations), there were 19.7% (43/218) children, which resulted in a suggestion to refer them to an early intervention center for a specific assessment by a professional. These types of suggestions were classified as alarms. That was the most severe system decision. Table 3 shows the results of the final assessments. The system has recommended repeating the remaining assessments in subsequent months to further evaluate and discard language issues. Suggestions that indicate repeat evaluation are classified as warning type.

Table 3. Distribution of Gades' final suggestionsa (N=218).

Description	Distribution, n (%)
Typical developmental	117 (53.7)
Refer to early childhood care	43 (19.7)
Repeat assessment (within 3 mo)	28 (12.8)
Repeat assessment (within 2 mo)	19 (8.7)
Repeat assessment (within 1 mo)	11 (5)

[a]More than half of the sample (n=117, 53.7%) were identified as typical language development. However, 19.7% (43/218) of assessments identified language difficulties as indicative of possible simple language delay or developmental language disorder.

Table 4 displays the distribution of assessment types by gender. It reveals that more alarm cases were identified in boys (29/218, 13.3%) compared to girls (14/218, 6.4%). In addition, both genders had a similar frequency of warning assessments and neurotypical results.

Table 4. Distribution of evaluations by suggestion type and gender. More alarm cases were identified in boys (29/218, 13.3%) compared with girls (14/218, 6.4%). Both genders had a similar rate of warning assessments and neurotypical results (N=218).

Gender and evaluation	Evaluations, n (%)
Female	
Alarm	14 (6.4)
Warning	29 (13.3)
No findings	56 (25.9)
Male	
Alarm	29 (13.3)
Warning	29 (13.3)
No findings	61 (27.9)

Gades identified more cases of alarm and warning in children already undergoing evaluation by specialists than from the center (alarm: 29.7%; warning: 32.1%), but we do not know whether these children had language disorders. Hence, among children already experiencing some difficulties, it identified that there may be some indication of language difficulties (ie, DLD or SLD).

Table 5 shows the distribution of evaluations that were classified as an alarm because it was detected that the assessed child did not meet the language acquisition milestones expected for their age. The incidence of language impairment from center 2 was comparable between boys and girls (15:10) and 30% (25/84) of the cases were detected as alarms. Meanwhile, for children without a suspicion of NDD coming from nursery school (center 1), difficulties in language and communication were detected in 13.4% (18/134) of the cases. However, a notable discrepancy was observed between boys and girls (14:4).

Table 5. Distributions of alarm evaluations by gender, month, and center[a].

Month	9	10	11	14	15	16	18	19	20	24	25	26	27	28	30	32	37	Total
Children center 1																		
Female	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	4
Male	0	1	0	1	0	0	4	0	0	1	1	2	1	0	1	1	1	14
Total	0	2	1	1	0	0	4	0	1	2	1	2	1	0	1	1	1	18
Children center 2																		
Female	1	0	0	0	2	0	1	3	0	0	0	1	1	1	0	0	0	10
Male	0	0	0	0	3	1	4	2	2	2	1	0	0	0	0	0	0	15
Total	1	0	0	0	5	1	5	5	2	2	1	1	1	1	0	0	0	25

[a]For children coming from center 1, language issues were detected in 13.4% (18/134) of the cases. The incidence of language issues from center 2 was comparable between male and female participants (15:10), with alarms detected in 30% (25/84) of the cases.

Notably, most of the suspected cases were detected in months 15, 18, 19 and 24. All these months correspond to the second year, encompassing children aged between 1 and 2 years. Upon thorough analysis of the evaluations during these months,

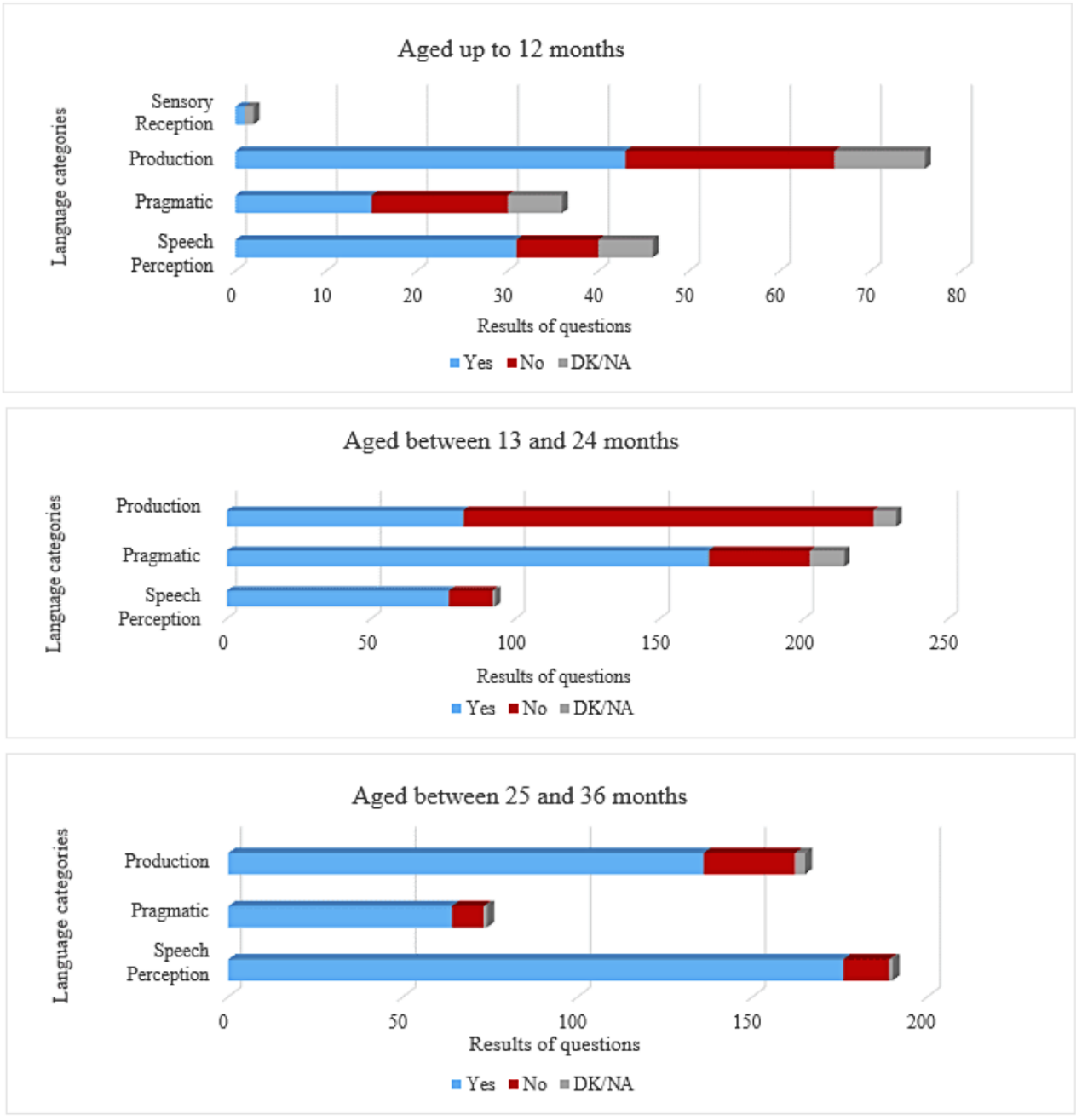
difficulties were discerned in the language production category for month 15. Half (23/40, 50%) of the children resulted in alarms. When compared to assessments from previous months, several observations emerged. First, most questions in this category were not been asked in previous months, as the child's age renders them too young to possess these competencies. Second, only 1 question, "first disyllables (one word in addition to dada and mama)?" was repeated in months 12 and 14, with half (8/14, 57%) of the assessed children answering negatively in both months. For month 15, all respondents answered negatively.

Month 18 had the highest number of detected suspected cases (9/11, 82%). During one of the meetings, educators expressed that they had observed more issues in language production, citing example, such as the child's inability to articulate certain words. In addition, similar questions asked in previous months also resulted in alarms. For instance, the one with the most error answers was "first bisyllables (two words in addition to dada and mama)?" In subsequent months, specifically month 19, identical pragmatics and language production questions were asked, and negative responses were received.

In month 24, 40% (4/10) of the evaluations raised alarms. Upon analyzing the questions, 3 were related to pragmatics and language comprehension, which were previously asked in earlier months. Most children satisfactorily answered these questions, indicating alignment with their current developmental stage. Therefore, the cases prompting alarms may signify potential instances of children with SLD or DLD.

Figure 4 depicts the results of questions categorized by language development areas. The sensory reception category was evaluated only for children aged up to 12 months to discard hearing issues. Children aged <12 months encountered more challenges in the production and pragmatics categories. Furthermore, there were no cases of hearing problems. The pragmatics category showed an equal balance between successes and failures. Conversely, children in the second year demonstrated notable difficulty in answering questions within the language production category, with a higher-than-average rate of incorrect responses. Other categories showed satisfactory performance.

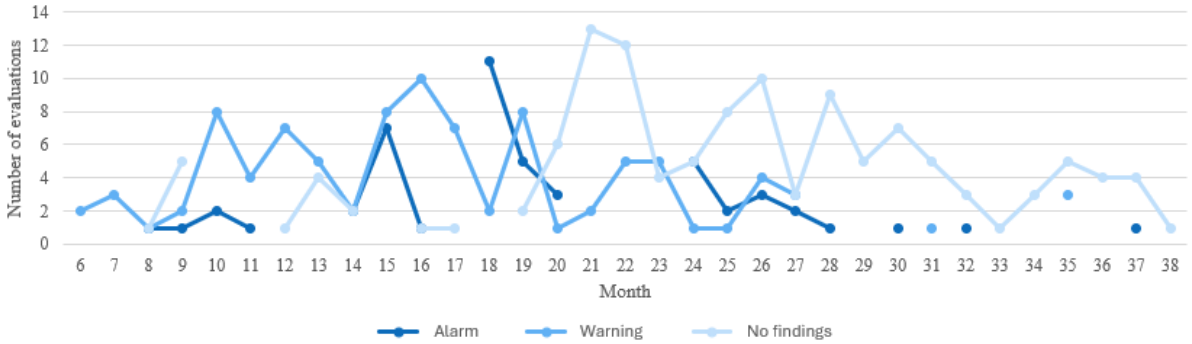
Figure 4. Distribution of results based on language-category questions. Each question is categorized according to language category (refer to Table 1 for more information) and can be answered with a yes, no, or do not know/no answer (DK/NA). Children aged from 13 to 24 months demonstrated notable difficulty in production questions. Children aged >24 months exhibited positive responses across most categories.



Finally, children aged >24 months exhibited predominantly positive responses across most categories. Hence, it becomes evident that children aged <24 months exhibit a higher frequency of errors, attributable to the inherent variability in language acquisition during this developmental stage. Consequently, there is a pressing need for greater flexibility in language milestones and extending their coverage across months.

As illustrated in Figure 5, this distribution reveals that instances of children with difficulty in language development are prevalent between months 15 and 27, gradually decreasing in the final months until 3 years of age.

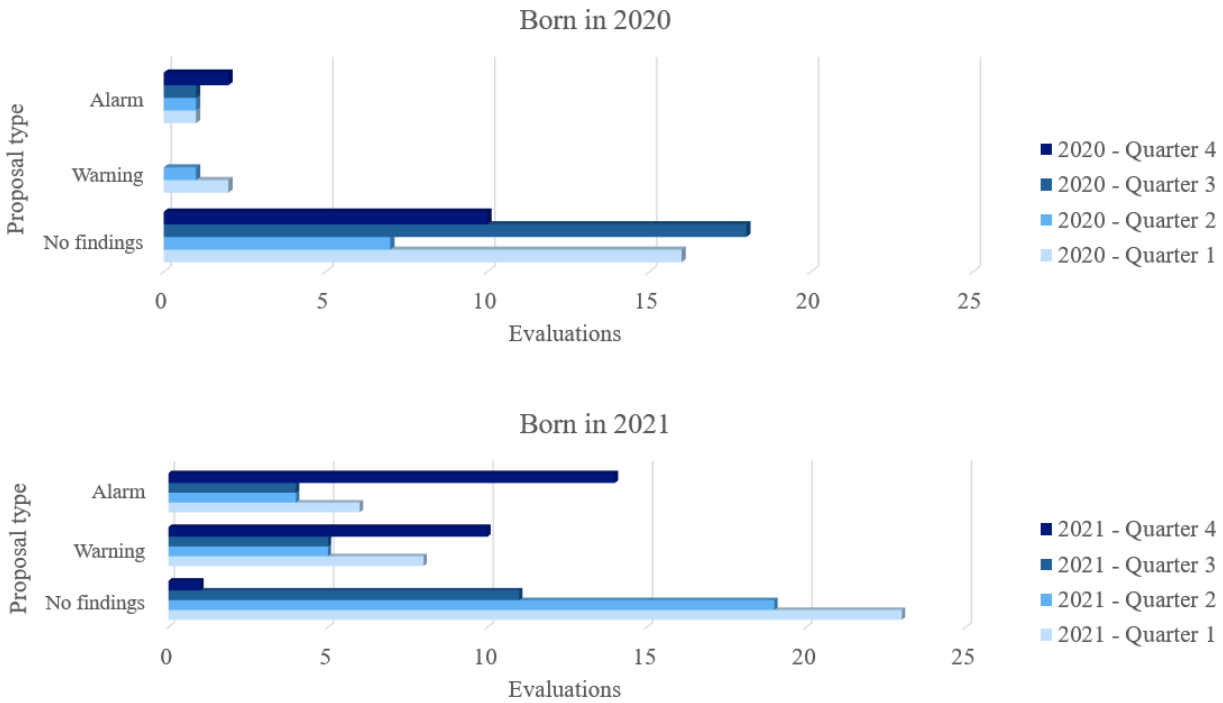
Figure 5. Distribution of suggestion type per month. Cases of children with language development issues were prevalent between months 15 and 27 (dark blue). The graphic reveals a gradual decline during the final months, with more typical development cases (pale blue).



In contrast, the evaluations were analyzed considering the control questions asked and the correlation between language development and the child's date of birth. In Madrid's nursery school classes, children were grouped based on their birth year. Consequently, at the beginning of the school year in September, a significant age gap existed between children born in the first quarter and those born in the last months. The latter group comprised the youngest in the class, resulting in a notable developmental difference between them.

The assessment outcomes for the youngest children in each class are presented in Figure 6. The youngest children were defined as those born from October to December 2020 for the age range of 25 to 36 months and from October to December 2021 for the age range of 13 to 24 months. This categorization was not applicable for children aged <12 months, as we did not have children from the last trimester. Hence, upon comparing the alarm for children born in 2020 and those born in 2021, it consistently appeared higher for those in the previous quarter. The youngest children tend to yield poorer results using Gades, as language development differs.

Figure 6. Assessment results by school quarter and suggestion type. For children born in 2020 (age between 25 and 36 months) alarm cases were more prevalent among children from quarter 4. However, these results are tempered by many more typical development cases at the same age. By contrast, for children born in 2021, the youngest children (quarter 4) tended to have alarm suggestions using the Gades platform.



The results of the assessment outcomes in relation to the control questions can be observed in Table 6. It is noteworthy that the 2 children who disclosed a family history of other NDD in the control questions were identified as cases with difficulties in language development (ie, SLD or DLD cases). Among cases with reported issues during pregnancy (ie, perinatal, prenatal, or premature), a notable percentage resulted in difficulties in language development (9/13, 69% including warnings). In the case of preterm children, only 8 (3.7%) of the 218 cases were identified as cases with difficulties in language development. Only one of these preterm cases, a girl, was assessed as an alarm. The 3 cases of preterm boys without previous NDD

demonstrated typical development. The remaining alarms in boys corresponded to cases of prenatal (3/218, 1.4%) and perinatal (4/218, 1.8%) risks. A similar scenario extended to bilingual cases, where only 1 child exhibited an alarm of language development difficulties.

Table 6. The child's evaluations were distributed by control information, gender, and suggestion typea (N=218).

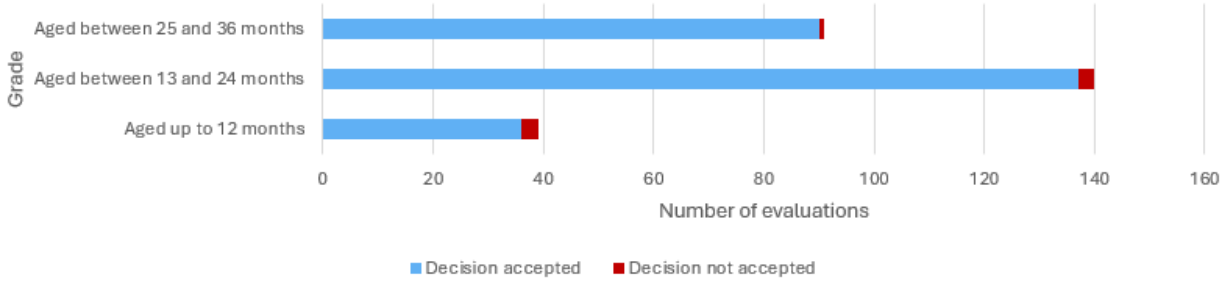
Gender and evaluation	Birth complications, n (%)		Family history, n (%)		Bilingual cases, n (%)	
	Center 1	Center 2	Center 1	Center 2	Center 1	Center 2
Female						
Alarm	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Warning	3 (1.4)	0 (0)	0 (0)	0 (0)	3 (1.4)	0 (0)
Typical development	1 (0.5)	0 (0)	0 (0)	0 (0)	4 (1.8)	0 (0)
Male						
Alarm	4 (1.8)	2 (0.9)	1 (0.5)	1 (0.5)	1 (0.5)	0 (0)
Warning	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Typical development	3 (1.4)	0 (0)	0 (0)	0 (0)	2 (0.9)	0 (0)

[a]Among children with birth complications, 69% (9/13) of them showed language development difficulties. Those with a family history of neurodevelopmental disorder also faced challenges, while only 1 bilingual child showed an alarm.

Gades' Performance

Gades' performance is assessed by the extent to which the system's decisions align with the perspectives of evaluators who support its recommendations. If there was disagreement with a recommendation, evaluators were prompted to justify dissent. Evaluators based their agreement or disagreement with the system's decision on their understanding of the child, considering factors such as age, family environment, and communicative intent. According to the results, Gades exhibited an accuracy rate of 97.4% (263/270). Out of 270 evaluations, only 7 (2.6%) suggestions were declined by evaluators. Figure 7 shows a higher disagreement rate with the system's decisions in children aged <24 months. Specifically, the accuracy rate was 92.3% up to 12 months, compared to 97.8% in the second grade (aged between 13 and 24 months) and 98.9% in the third grade (aged between 25 and 36 months).

Figure 7. Gades acceptance per grade. Out of 270 evaluations, only 7 proposals were declined (ie, 3 for evaluations of the first and second grades and 1 for evaluations in children aged >24 months). Even so, the acceptance of Gades remains above 90% in all courses.



Considering the total of 270 assessments conducted, including those where children were evaluated twice, the accuracy of Gades' decisions is depicted in Table 7. Among the suggestions provided by the Gades system, where evaluators disagreed (7/270, 2.6%), some suggested referring the child to early intervention for a specialized assessment (5/270, 1.9%).

Table 7. Gades' accuracy by the suggestion typea.

Suggestion description	Accepted suggestion, n (%)	
	No (n=7)	Yes (n=263)
Refer to early childhood care	5 (1.9)	45 (16.7)
Typical development	0 (0)	125 (46.3)
Repeat assessment (within 2 mo)	2 (0.74)	29 (10.7)
Repeat assessment (within 3 mo)	0 (0)	44 (16.3)
Repeat assessment (within 1 mo)	0 (0)	20 (7.4)

[a]Evaluators did not agree with Gades' suggestion in 2.6% (7/270) of evaluations. Most of these unaccepted suggestions were for the most severe cases identified by Gades ("refer to early childhood care"). Gades exhibit an accuracy rate of 97.41%.

When the evaluator disagreed with the Gades' decision, they were required to provide a comment stating the reason for nonacceptance. Thus, several reasons explain why the evaluators did not accept Gades' decision (Table 8). In 4 instances (ie, for months 10, 12, 15, and 18), they noted that the questions were not tailored to the child's developmental stage. In 2 cases from 18 months, the evaluator rejected the suggestion. Moreover, in case "The questions do not align with the child's developmental stage," the assessor disagreed with the suggestion and conducted an immediate reassessment. This subsequent assessment issued a warning (ie, to repeat the evaluation in the next months). In addition, in 2 cases involving 2 boys, the evaluators intended to continue observing the child's progress as they had identified other communicative skills. Finally, there was no disagreement for typical development cases.

Table 8. Evaluator's reasons for rejecting the final suggestion from Gadesa.

Reason	Children's gender	Evaluation month
Refer to early childhood care		
"We will monitor the situation closely. If necessary, we will consult the Early Assessment Team for guidance."	Male	26
"There are alternative forms of communication besides verbal language, and he demonstrates proficiency in them. We will observe his development and see how it progresses."	Male	18
"The questions do not align with the child's developmental stage."	Female	10
"The questions do not align with the child's developmental stage."	Male	10
"The questions do not align with the child's developmental stage."	Female	15
Repeat assessment (within 2 mo)		
"The questions do not align with the child's developmental stage."	Female	12
"We consider that it is not necessary at this stage."	Male	18

[a]The suggestions that were not accepted relate to "refer to early childhood care and repeat assessment (within 2 months)". Most of the reasons were that the assessment conducted by Gades was not aligned with the child's developmental stage.

Overview of the Obtained Functional Results

Numerous interviews were conducted with the evaluation staff, and they completed a web-based form assessing the usefulness and practicality of the Gades tool in uncontrolled educational settings, following the guidelines outlined in the referenced study [41]. None of the evaluators had previous experience using this specific platform. It was noted that preschool educators and professionals in early childhood educational and psychopedagogical centers found Gades highly acceptable. Most (18/24, 75%) of the evaluators found the system easy to use, although it required dedicated time. They experienced occasional technological issues such as student search processes or time sessions. A quarter felt previous technological knowledge was necessary.

The educators used an observational methodology during the school day to respond to the form's questions, considering it noninvasive and conducive to respectful support within an appropriate environment for the child. They suggested that implementing the Gades assessment during the first term of the school year, particularly during the adaptation period, could be challenging due to their limited familiarity with the child. In addition, they proposed conducting evaluations every 3 months for cases initially indicating "typical development" to ensure ongoing progress.

While control questions were valuable for accurate results from the Gades system, access to certain information, particularly related to birth difficulties, was limited for many children. Regarding the comprehension and appropriateness of the KB questions, educators encountered comprehension difficulties in only 2 out of the 108 questions, which they resolved by consulting the Gades' technical staff. They noted that some questions were unsuitable for the child's developmental stage. They highlighted the absence of questions about family context, communication dynamics, and disruptive behaviors in the second and third years.

Regarding recommendations, 75% (18/24) of the evaluators reported that they would suggest Gades to other nursery schools, emphasizing the need for previous DLD training to observe children effectively. They reported high knowledge acquisition and insight into their students' language development. For instance, they had observed that girls aged <12 months initiated communication earlier and more fluently.

Discussion

Principal Findings

This study's core findings underscore the Gades platform's effectiveness as a robust tool for screening language disorders in preschool environments, successfully identifying language difficulties in 19.7% (43/218) of the children assessed. This rate, with a notable prevalence in boys (29/218, 13.3%) compared to girls (14/218, 6.4%), is significantly higher than the traditional prevalence estimates of 2% to 7% for DLDs [8,17,21,23,43]. This discrepancy suggests that Gades may be particularly sensitive to early signs of language difficulties or may also capture cases of SLD. As a screening tool rather than a diagnostic system, Gades aims to identify children at higher risk for DLD or SLD, which might lead to elevated detection rates compared to studies focused strictly on diagnosis. Comprehensive assessments by specialists at early intervention

centers are required for a definitive diagnosis, highlighting the necessity of further evaluation following initial screening outcomes.

The study validated Gades' technical and functional capabilities in uncontrolled educational settings without direct intervention by an NSLT. The platform's utility in guiding nonexpert users, such as preschool teachers, through the screening process was particularly noteworthy. Teachers could use Gades effectively to observe and assess language development, reflecting its potential to empower educators with limited specialized training in language disorders.

Moreover, Gades demonstrated a high accuracy rate (97.41%) in aligning with the educators' assessments, underscoring its reliability and the robustness of its underlying algorithms and KB. The discrepancies between the system's recommendations and the educators' judgments were minimal, indicating a high level of agreement and trust in the platform's diagnostic suggestions.

The findings also illuminated specific age ranges (between 15 and 27 months) where language difficulties are most prevalent, reinforcing the importance of targeted early intervention during this critical developmental period. The ability of Gades to adapt its screening approach based on the child's age and specific educational context was a key factor in its effectiveness.

These promising results suggest that integrating Gades into regular preschool assessments could significantly enhance the early detection of language disorders, potentially leading to more timely and effective interventions.

Extending Detection to Natural Environments

This study highlights the importance of checking for developmental disorders similar to DLD in natural settings, such as preschools, where children naturally spend much time. DLD, similar to autism and ADHD, is among the most common developmental challenges but often goes unrecognized because its symptoms may resemble simple speech delays or overlap with other developmental disorders, leading to delayed diagnoses. Equipping educators, teachers, and school professionals with the right knowledge and support is crucial because they can spot early signs of these disorders. Early detection, significantly enhanced through collaboration with a multidisciplinary team, is key to improving a child's prospects for developing language skills effectively.

Applicability and Scalability

The Gades platform is designed to be versatile and scalable, making it a valuable tool for educators worldwide, especially in regions with limited access to specialized language disorder services. This flexibility is crucial for adapting to diverse linguistic and cultural environments across global educational systems. To effectively implement Gades in these diverse settings, developing training programs sensitive to local languages and educational standards is essential. These programs should equip educators with the skills to use the platform effectively and to recognize early signs of language disorders, even without specialized training.

Considering the technological limitations in different regions, Gades should be designed to operate in low-bandwidth environments and comply with local data protection regulations. This will make the platform accessible and trustworthy. Enhancing the platform's database to support multiple languages and reflect cultural specifics can make Gades a more inclusive tool that accurately reflects diverse child development patterns.

Practical implementation strategies include collaborating with local educational authorities to incorporate Gades into routine screening processes. This might involve tailoring the platform to meet local developmental benchmarks and language nuances. Partnering with local educational institutions can provide continuous support and feedback, helping to refine the platform.

Gades could offer versions that range from sophisticated, high-tech applications to more basic, web-accessible formats to accommodate varying technological capabilities. Establishing a community of practice among Gades users would encourage sharing best practices and provide essential peer support, enhancing the platform's overall effectiveness and user experience.

Limitations

One significant limitation of the Gades system is its reliance on the accuracy of the input provided by educators, who may not have specialized training in identifying language disorders. This dependency can introduce biases or inaccuracies in the screening process, as nonexpert interpretations of language development may vary widely, affecting the reliability of the outcomes. To mitigate these potential biases, it is essential to implement comprehensive training programs for educators using the platform. These programs should focus on familiarizing them with common language development milestones and the specific signs of language disorders. In addition, incorporating automated prompts and guidelines within the Gades system can help standardize recorded observations, reducing subjective interpretation errors.

Further refining the interface and feedback mechanisms of Gades can also enhance the accuracy of the data collected. For instance, the platform could include illustrative examples or short training videos on typical versus atypical language behaviors to guide educators' assessments before they input data. Regular updates and calibrations based on user feedback and new research in language development could further improve the system's precision and adaptability to real-world educational environments.

The study's limitations extend beyond the technical aspects of the Gades platform to include concerns about the sample size and the uncontrolled implementation settings. The small sample size may not adequately represent the broader population of preschool children, limiting the generalizability of the findings. In addition, while providing real-world insights into the use of Gades, the uncontrolled setting might introduce variability in how the platform is used across different environments, potentially influencing the consistency and applicability of the results. These factors could lead to higher variability in outcomes, which might not fully indicate the platform's efficacy under different or more controlled conditions.

Future Directions

Future research should include larger and more diverse populations to address the limitations mentioned to ensure that findings are robust and widely applicable. Controlled trials, where variables can be more tightly managed, would also help understand the platform's effectiveness across varied educational settings and populations. This would allow for more reliable validation of the platform's capabilities and identify specific conditions under which it performs best, guiding more targeted improvements.

To enhance Gades' effectiveness and reliability, it is imperative to refine the tool by expanding the KB to better differentiate between DLD and SLD at earlier stages. Adapting the system to align more closely with the actual ages of children rather than their academic year will provide more accurate assessments. Implementing educator training modules will equip them to effectively recognize early signs of language difficulties. Future versions of Gades should include a nuanced approach for assessing the youngest children in the class, considering their specific developmental stages and potential delays without prematurely recommending specialist evaluations. These improvements will ensure that Gades supports early detection and integrates smoothly into regular preschool assessments, potentially enhancing the effectiveness of interventions for language disorders.

Conclusions

This study confirms that DLD remains a commonly overlooked condition with significant variability and potential overlaps with other developmental disorders. Importantly, our findings reinforce the critical nature of age-specific screenings, with the most alarms for language difficulties arising in second-year grades, supporting the hypothesis that this period is crucial for detecting DLD risk factors. Despite the conservative and generic screening process, which identified language difficulties in 19% of cases—higher than the 7% prevalence typically noted in other studies—some of these instances may represent SLD. This suggests a need for refining language milestones within the Gades system to more effectively distinguish between DLD and SLD.

The qualitative outcomes from the Gades platform demonstrate its value in designing tailored educational programs and therapeutic interventions. By integrating such tools without the need for direct support from NSLT, Gades facilitates a deeper understanding of individual language development patterns. This insight enables educators, particularly those without specialized training in language disorders, to identify and address developmental challenges more effectively and earlier. The platform has been well-received in terms of usability and practicality, enhancing educators' ability to actively monitor and support language development.

Moreover, implementing Gades has provided substantial learning opportunities for preschool teachers. Educators with no previous expertise in DLD have gained significant knowledge and skills in informal screening processes, which has had a transformative impact on their ability to recognize and respond to language development issues. This empowerment of teachers underscores the potential of Gades to serve as a foundational tool in early education settings, enhancing early detection and intervention strategies.

To extend the utility of Gades beyond this study context, it is crucial to provide practical recommendations for its implementation in various educational settings. Tailoring the platform to accommodate different regional educational standards and linguistic backgrounds can maximize its effectiveness and reach. In addition, the successful deployment of Gades is contingent upon comprehensive training programs for educators. These programs should focus on enhancing understanding of language milestones, screening techniques, and the specific functionalities of the Gades platform. Effective training could be implemented through web-based modules, workshops, and ongoing support systems to ensure educators can proficiently use the tool.

In conclusion, the Gades platform represents a significant advancement in the field of educational technology for screening language disorders. By facilitating detailed observation and reporting of language development, Gades supports educators in their daily interactions with children and contributes to a broader strategy for addressing DLDs in early childhood education. The adoption of such tools, accompanied by adequate training and tailored implementation strategies, holds the promise of significantly improving outcomes for children at risk of DLD and SLD across diverse educational landscapes.

Acknowledgments

The authors would like to thank the team of educators and preschool teachers from Escuela Infantil Sabina and Patricia Calcerrada Muñoz from Equipo de Orientación Educativa y Psicopedagógica de Atención Temprana de Parla (Madrid, Spain) for their collaboration in performing the experiment presented here. In addition, the authors give special thanks to Esther González Dávila and Elena Jambrina Prieto from the Asociación del Trastorno Específico del Lenguaje de Madrid for their support in this study.

Data Availability

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

Authors' Contributions

MD-P and MLM-R were responsible for conceptualization. MD-P and AMG-P were responsible for data curation, formal analysis, and methodology. SB-G was responsible for funding acquisition. MD-P, DLG, and MLM-R were responsible for investigation.

MLM-R was responsible for resources. MLM-R and IPDLC were responsible for project administration. MD-P and DLG were responsible for software. MLM-R and IPDLC were responsible for supervision. MD-P, MLM-R, and AMG-P were responsible for validation. MD-P and SB-G were responsible for visualization. MD-P, SB-G, IPDLC, and MLM-R were responsible for writing the original draft and review and editing.

Conflicts of Interest

None declared.

References

1. Kidd E, Donnelly S. Individual differences in first language acquisition. Annu Rev Linguist 2020 Jan 14;6(1):319-340. [doi: 10.1146/annurev-linguistics-011619-030326]
2. Kuhl PK. Early language acquisition: cracking the speech code. Nat Rev Neurosci 2004 Nov;5(11):831-843. [doi: 10.1038/nrn1533] [Medline: 15496861]
3. Nayeb L, Lagerberg D, Westerlund M, Sarkadi A, Lucas S, Eriksson M. Modifying a language screening tool for three-year-old children identified severe language disorders six months earlier. Acta Paediatr 2019 Sep 25;108(9):1642-1648. [doi: 10.1111/apa.14790] [Medline: 30896050]
4. Sansavini A, Favilla ME, Guasti MT, Marini A, Millepiedi S, Di Martino MV, et al. Developmental language disorder: early predictors, age for the diagnosis, and diagnostic tools. A scoping review. Brain Sci 2021 May 17;11(5):654 [FREE Full text] [doi: 10.3390/brainsci11050654] [Medline: 34067874]
5. Alás Rupérez A, Ramos Sánchez I, Machado Casas IS, Martín Fernández-Mayoralas D, Gortázar Díaz M, Aguilera Albesa S. Trastornos del lenguaje, del habla y de la comunicación. Conceptos, clasificación y clínica. Protoc Diagn Pediatr 2022;1:19-30 [FREE Full text]
6. Rowe ML. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. Child Dev 2012 Jun 20;83(5):1762-1774 [FREE Full text] [doi: 10.1111/j.1467-8624.2012.01805.x] [Medline: 22716950]
7. Asorey MJ, Fernández MP. Dificultades de Aprendizaje y Trastornos del Desarrollo: Manual Didáctico. Madrid, Spain: Ediciones Pirámide; 2014.
8. Tomblin JB, Zhang X, Buckwalter P, O'Brien M. The stability of primary language disorder: four years after kindergarten diagnosis. J Speech Lang Hear Res 2003 Dec;46(6):1283-1296. [doi: 10.1044/1092-4388(2003/100)] [Medline: 14700355]
9. Tomblin JB, Records NL, Buckwalter P, Zhang X, Smith E, O'Brien M. Prevalence of specific language impairment in kindergarten children. J Speech Lang Hear Res 1997 Dec;40(6):1245-1260 [FREE Full text] [doi: 10.1044/jslhr.4006.1245] [Medline: 9430746]
10. Clegg J, Hollis C, Mawhood L, Rutter M. Developmental language disorders--a follow-up in later adult life. Cognitive, language and psychosocial outcomes. J Child Psychol Psychiatry 2005 Feb;46(2):128-149. [doi: 10.1111/j.1469-7610.2004.00342.x] [Medline: 15679523]
11. Fradley K, Oldfield J, Marshall J, Toseeb U. Early life risk factors for adolescent mental health difficulties for individuals at risk of developmental language disorder. JORSEN 2024 Mar 22;24(3):713-728. [doi: 10.1111/1471-3802.12654]
12. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013.
13. Lara EM. Trastorno Específico Del Lenguaje (TEL): Avances en El Estudio De Un Trastorno Invisible. Madrid, Spain: Ediciones Pirámide; 2016.
14. Bichay-Awadalla K, Qi CH, Bulotsky-Shearer RJ, Carta JJ. Bidirectional relationship between language skills and behavior problems in preschool children from low-income families. J Emot Behav Disord 2019 Jun 06;28(2):114-128. [doi: 10.1177/1063426619853535]
15. Romeo RR, Leonard JA, Robinson ST, West MR, Mackey AP, Rowe ML, et al. Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. Psychol Sci 2018 May 14;29(5):700-710 [FREE Full text] [doi: 10.1177/0956797617742725] [Medline: 29442613]
16. Kohnert K, Windsor J, Ebert KD. Primary or "specific" language impairment and children learning a second language. Brain Lang 2009 May;109(2-3):101-111 [FREE Full text] [doi: 10.1016/j.bandl.2008.01.009] [Medline: 18313136]
17. Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. J Child Psychol Psychiatry 2016 Nov;57(11):1247-1257 [FREE Full text] [doi: 10.1111/jcpp.12573] [Medline: 27184709]
18. López IJ, Förster J. Trastornos del neurodesarrollo: dónde estamos hoy y hacia dónde nos dirigimos. Revista Médica Clínica Las Condes 2022 Jul;33(4):367-378. [doi: 10.1016/j.rmclc.2022.06.004]
19. Francés L, Quintero J, Fernández A, Ruiz A, Caules J, Fillon G, et al. Current state of knowledge on the prevalence of neurodevelopmental disorders in childhood according to the DSM-5: a systematic review in accordance with the PRISMA criteria. Child Adolesc Psychiatry Ment Health 2022 Mar 31;16(1):27 [FREE Full text] [doi: 10.1186/s13034-022-00462-1] [Medline: 35361232]

20. Auza BA, Murata C, Peñaloza C. "Early detection of Spanish-speaking children with developmental language disorders: concurrent validity of a short questionnaire and a screening test". J Commun Disord 2023;104:106339. [doi: 10.1016/j.jcomdis.2023.106339] [Medline: 37247522]
21. Bosch R, Pagerols M, Rivas C, Sixto L, Bricollé L, Español-Martín G, et al. Neurodevelopmental disorders among Spanish school-age children: prevalence and sociodemographic correlates. Psychol Med 2021 Jan 13;52(14):3062-3072. [doi: 10.1017/s0033291720005115]
22. Carballal Mariño M, Gago Ageitos A, Ares Alvarez J, del Rio Garma M, García Cendón C, Goicoechea Castaño A, et al. Prevalence of neurodevelopmental, behavioural and learning disorders in paediatric primary care. Anales de Pediatría (English Edition) 2018 Sep;89(3):153-161. [doi: 10.1016/j.anpede.2017.10.005]
23. Calder SD, Brennan-Jones CG, Robinson M, Whitehouse A, Hill E. The prevalence of and potential risk factors for Developmental Language Disorder at 10 years in the Raine Study. J Paediatr Child Health 2022 Nov 03;58(11):2044-2050 [FREE Full text] [doi: 10.1111/jpc.16149] [Medline: 35922883]
24. Girolamo TM, Rice ML, Selin CM, Wang CJ. Teacher educational decision making for children with specific language impairment. Am J Speech Lang Pathol 2022 May 10;31(3):1221-1243. [doi: 10.1044/2021_ajslp-20-00366]
25. Bloder T, Eikerling M, Rinker T, Lorusso ML. Speech and language therapy service for multilingual children: attitudes and approaches across four European countries. Sustainability 2021 Nov 03;13(21):12143. [doi: 10.3390/su132112143]
26. Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, CATALISE Consortium. CATALISE: a multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. PLoS One 2016;11(7):e0158753 [FREE Full text] [doi: 10.1371/journal.pone.0158753] [Medline: 27392128]
27. Waine H, Bates S, Frizelle P, Oh TM. UK speech and language therapists' assessment of children's expressive language, and functional impairment and impact, following the CATALISE publications. Int J Lang Commun Disord 2023 Apr 27;58(5):1570-1587. [doi: 10.1111/1460-6984.12883] [Medline: 37102426]
28. Díaz MG, Sánchez IR, Casas IS, Rupérez AA, Fernández-Mayoralas DM, Aguilera Albesa S, et al. Trastornos del lenguaje, del habla y de la comunicación: desarrollo del lenguaje, signos de alerta, enfoque diagnóstico y tratamiento. Protoc Diagn ter Pediatr 2022;1:31-42 [FREE Full text] [doi: 10.14201/gredos.83174]
29. Gabaldón-Pérez AM, Martín-Ruiz ML, Díez-Muñoz F, Díez-Muñoz M, Máximo-Bocanegra N, Pau de la Cruz I. The potential of digital screening tools for childhood ADHD in school environments: a preliminary study. Healthcare 2023;11(20):2795. [doi: 10.3390/healthcare11202795]
30. Eikerling M, Lorusso M. A web-platform for DLD screening in Italian-German-speaking children: preliminary data on concurrent and predictive validity. Lingue e Linguaggi 2022;21(1):193-212 [FREE Full text] [doi: 10.1418/104455]
31. Eikerling MR, Bloder TS, Lorusso ML. A nonword repetition task discriminates typically developing Italian-German bilingual children from bilingual children with developmental language disorder: the role of language-specific and language-non-specific nonwords. Front Psychol 2022;13:826540 [FREE Full text] [doi: 10.3389/fpsyg.2022.826540] [Medline: 35719570]
32. Eikerling M, Andreoletti M, Secco M, Luculli B, Cha G, Castro S, et al. Remote screening for developmental language disorder in bilingual children: preliminary validation in Spanish–Italian speaking preschool children. Appl Sci 2023 Jan 21;13(3):1442. [doi: 10.3390/app13031442]
33. Equipos de orientación educativa y psicopedagógica. Comunidad de Madrid. URL: https://www.comunidad.madrid/servicios/educacion/equipos-orientacion-educativa-psicopedagogica#panel-135799 [accessed 2024-02-14]
34. Fiestas CE, Lugo-Neris MJ, Pratt AS, Peña ED, Bedore LM. Spanish language and literacy intervention for bilingual children at risk for developmental language disorder: a pilot study. Top Lang Disord 2021;41(4):309-321. [doi: 10.1097/TLD.0000000000000263]
35. Martín Ruiz ML, Valero Duboy M, Torcal Loriente C, Pau de la Cruz I. Evaluating a web-based clinical decision support system for language disorders screening in a nursery school. J Med Internet Res 2014 May 28;16(5):e139 [FREE Full text] [doi: 10.2196/jmir.3263] [Medline: 24870413]
36. Bowman R, Cooney O, Thieme A, Cowan BR, Doherty G. Pervasive therapy: designing conversation-based interfaces for ecological momentary intervention. IEEE Pervasive Comput 2022 Oct 1;21(4):55-63. [doi: 10.1109/mprv.2022.3209068]
37. Berrezueta-Guzman J, Martin-Ruiz ML, Pau I, Krusche S. A user-centered methodology approach for the development of robotic assistants for pervasive unsupervised occupational therapy. In: Proceedings of the 8th International Conference on Robotics and Artificial Intelligence. 2022 Presented at: ICRAI '22; November 18-20, 2022; Singapore, Singapore. [doi: 10.1145/3573910.3573925]
38. Robles-Bykbaev V, López-Nores M, García-Duque J, Pazos-Arias JJ, Arévalo-Lucero D. Evaluation of an expert system for the generation of speech and language therapy plans. JMIR Med Inform 2016 Jul 01;4(3):e23 [FREE Full text] [doi: 10.2196/medinform.5660] [Medline: 27370070]
39. Toki EI, Zakopoulou V, Tatsis G, Plachouras K, Siafaka V, Kosma EI, et al. A game-based smart system identifying developmental speech and language disorders in child communication: a protocol towards digital clinical diagnostic procedures. In: Proceedings of the Interactive Mobile Communication, Technologies and Learning. 2021 Presented at: IMCL 2021; November 4-5, 2021; Virtual Event. [doi: 10.1007/978-3-030-96296-8_50]

40. Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, the CATALISE-2 Consortium. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. J Child Psychol Psychiatry 2017 Oct;58(10):1068-1080 [FREE Full text] [doi: 10.1111/jcpp.12721] [Medline: 28369935]
41. Sevilla-Gonzalez MD, Moreno Loaeza L, Lazaro-Carrera LS, Bourguet Ramirez B, Vázquez Rodríguez A, Peralta-Pedrero ML, et al. Spanish version of the system usability scale for the assessment of electronic tools: development and validation. JMIR Hum Factors 2020 Dec 16;7(4):e21161 [FREE Full text] [doi: 10.2196/21161] [Medline: 33325828]
42. Fernald A, Marchman VA, Weisleder A. SES differences in language processing skill and vocabulary are evident at 18 months. Dev Sci 2013 Mar 08;16(2):234-248 [FREE Full text] [doi: 10.1111/desc.12019] [Medline: 23432833]
43. Nayeb L, Lagerberg D, Sarkadi A, Salameh EK, Eriksson M. Identifying language disorder in bilingual children aged 2.5 years requires screening in both languages. Acta Paediatr 2021 Jan;110(1):265-272. [doi: 10.1111/apa.15343] [Medline: 32869381]
44. O'Toole C, Gatt DR. Risk and Protective Environmental Factors for Early Bilingual Language Acquisition. Milton Park, UK: Taylor & Francis; 2019.
45. What is a graph database. Neo4j. URL: https://neo4j.com/docs/getting-started/graph-database/ [accessed 2024-02-14]
46. Población residente por fecha, sexo y generación (edad a 31 de diciembre). Instituto Nacional de Estadística. URL: https://ine.es/jaxiT3/Datos.htm?t=56941 [accessed 2024-12-14]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
DLD: developmental language disorder
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
KB: knowledge base
NDD: neurodevelopmental disorder
NSLT: nursery school language therapist
SLD: simple language delay

Edited by C Jacob; submitted 10.05.24; peer-reviewed by A Cruz-Santos, MÁV Duboy, D Casado-Mansilla, R Melguizo, R Soto-Varela; comments to author 18.11.24; revised version received 20.11.24; accepted 20.01.25; published 14.03.25.

Please cite as:
Dolón-Poza M, Gabaldón-Pérez AM, Berrezueta-Guzman S, López Gracia D, Martín-Ruiz ML, Pau De La Cruz I
Enhancing Early Language Disorder Detection in Preschools: Evaluation and Future Directions for the Gades Platform
JMIR Hum Factors 2025;12:e60424
URL: https://humanfactors.jmir.org/2025/1/e60424
doi:10.2196/60424
PMID:

Factors Determining Acceptance of Internet of Things in Medical Education: Mixed Methods Study

Khadija Alhumaid[1*], PhD; Kevin Ayoubi[2*], PhD; Maha Khalifa[1*], PhD; Said Salloum[3*], PhD

[1]Research & Innovation Division, Rabdan Academy, Abu Dhabi, United Arab Emirates

[2]English Faculty, Higher Colleges of Technology – CERT, Al Ain, United Arab Emirates

[3]School of Computing, Skyline University College, University City, Sharjah, United Arab Emirates

[*]all authors contributed equally

Corresponding Author:
Said Salloum, PhD
School of Computing, Skyline University College, University City, Sharjah, United Arab Emirates

Abstract

Background: The global increase in the Internet of Things (IoT) adoption has sparked interest in its application within the educational sector, particularly in colleges and universities. Previous studies have often focused on individual attitudes toward IoT without considering a multiperspective approach and have overlooked the impact of IoT on the technology acceptance model outside the educational domain.

Objective: This study aims to bridge the research gap by investigating the factors influencing IoT adoption in educational settings, thereby enhancing the understanding of collaborative learning through technology. It seeks to elucidate how IoT can facilitate learning processes and technology acceptance among college and university students in the United Arab Emirates.

Methods: A questionnaire was distributed to students across various colleges and universities in the United Arab Emirates, garnering 463 participants. The data collected were analyzed using a hybrid approach that integrates structural equation modeling (SEM) and artificial neural network (ANN), along with importance-performance map analysis to evaluate the significance and performance of each factor affecting IoT adoption.

Results: The study, involving 463 participants, identifies 2 primary levels at which factors influence the intention to adopt IoT technologies. Initial influences include technology optimism (TOP), innovation, and learning motivation, crucial for application engagement. Advanced influences stem from technology acceptance model constructs, particularly perceived ease of use (PE) and perceived usefulness (PU), which directly enhance adoption intentions. Detailed statistical analysis using partial least squares–SEM reveals significant relationships: TOP and innovativeness impact PE (β=.412, *P*=.04; β=.608, *P*=.002, respectively), and PU significantly influences TOP (β=.381, *P*=.04), innovativeness (β=.557, *P*=.003), and learning motivation (β=.752, *P*<.001). These results support our hypotheses (H1, H2, H3, H4, and H5). Further, the intention to use IoT is significantly affected by PE and usefulness (β=.619, *P*<.001; β=.598, *P*<.001, respectively). ANN modeling enhances these findings, showing superior predictive power (R^2=89.7%) compared to partial least squares–SEM (R^2=86.3%), indicating a more effective identification of nonlinear associations. Importance-performance map analysis corroborates these results, demonstrating the importance and performance of PU as most critical, followed by technology innovativeness and optimism, in shaping behavioral intentions to use IoT.

Conclusions: This research contributes methodologically by leveraging deep ANN architecture to explore nonlinear relationships among factors influencing IoT adoption in education. The study underscores the importance of both intrinsic motivational factors and perceived technological attributes in fostering IoT adoption, offering insights for educational institutions considering IoT integration into their learning environments.

(JMIR Hum Factors 2025;12:e58377) doi:10.2196/58377

KEYWORDS

collaborative learning; student; college; university; education; Internet of Things; IoT; technology acceptance model; technology optimism; TAM; experience; attitude; opinion; perception; perspective; acceptance; adoption; survey; questionnaire; ANN; deep learning; structural equation modeling; neural network; intent; use; medical education; artificial neural network; technology innovation

Introduction

Background

The use of Internet of Things (IoT) applications has been implemented in medical education projects [1]. IoT has the ability to completely transform the educational landscape by offering a more adaptable and quantifiable educational system that unites teachers and students under a single technological roof [2]. IoT innovation plays a significant role in transforming training at all levels, from school and college to university education [3]. Everyone, including students, instructors, and college campuses, can benefit from this innovation. Educators and administrators can leverage the power of IoT to connect people with devices and data, enabling them to gain valuable insights that have not been previously used in education. The traditional human-centered educational system has been transformed into an IoT-based one [4-6] using IoT. IoT has been leveraged to change the conventional personalized schooling system to an IoT-based system [4-6]. Figure 1 illustrates a cohort of university students actively engaging with IoT technology within an academic context. These individuals are adorned with sophisticated wearable devices, including smart glasses and smartwatches, which project information into their visual field, indicative of a digital and augmented educational milieu. The students are depicted with a palpable focus, using a portable electronic apparatus, likely for the manipulation of or interaction with the data overlay provided by their IoT devices.

The IoT has the potential to transform institutional practices and enhance learning capabilities across various levels and domains. University lecturers, students, and support staff can leverage large IoT platforms successfully. However, there is room for improvement in the utilization of IoT technology across different educational institutions. Researchers, scholars, and students can collaborate to develop IoT systems, devices, applications, and services, leading to the evolution of the educational environment as an increasingly dynamic and globally relevant subject. Better IoT deployments in colleges and universities have significantly improved positively to the creation of efficient and useful educational resources [7,8].

Figure 1. Students using augmented reality and Internet of Things technology (generated by Openart AI).

This research aims to explore how IoT technology is adopted and used by educators and students in universities across the Middle East. By integrating the technology acceptance model (TAM) and additional external factors, the study seeks to evaluate the effectiveness of IoT as an educational tool from the perspectives of both students and educators. Using the TAM as a metric, along with external variables, the research intends to identify and analyze the factors that influence the acceptance and use of IoT in medical education. The analysis uses a hybrid framework that combines structural equation modeling (SEM) and artificial neural networks (ANN) to examine how intrinsic motivational factors and perceived technological attributes affect IoT adoption. The SEM-ANN approach was specifically chosen to leverage the strengths of both methodologies [9,10]. SEM is highly effective in assessing the relationships between observed and latent variables, providing clarity on both the direct and indirect effects within the hypothesized model [11]. This method allows for robust statistical analysis capabilities, making it ideal for hypothesis testing and understanding the structural relationships among the theoretical constructs [12]. Conversely, ANN is used for its superior ability to model complex nonlinear relationships between variables, which are often not adequately captured by traditional linear models like SEM. ANN's data-driven nature allows it to directly learn and adapt to these relationships from the data, thereby enhancing the model's predictive accuracy and robustness. By integrating SEM with ANN, the study not only validates the theoretical framework through SEM's rigorous statistical analysis but also enhances the predictive power and generalization capabilities of the model with ANN's computational intelligence. This hybrid approach is particularly effective in exploring deeper, nonlinear interactions within the data, offering a more comprehensive understanding of the factors influencing IoT adoption. Given the complexities of modern datasets, which often exhibit nonlinear and digital behaviors among variables, this methodological integration is well-suited to achieving the research objectives and provides a justified, robust approach for the study.

The proposed hypotheses are: (1) intrinsic motivational factors like technology optimism (TOP), innovation, and learning motivation (LMT) significantly impact students' intention to adopt IoT; (2) perceived ease of use (PE) and perceived usefulness (PU), fundamental components of the TAM, strongly predict IoT adoption intentions among students in the United Arab Emirates.

This paper is organized as follows: "Literature Review" section delves into the existing research related to the adoption of IoT technologies, highlighting key theories and previous findings that set the groundwork for this study. Section "Methodology" details the research design, sampling methods, data collection procedures, and analytical techniques used to investigate the hypotheses. Section "Findings" presents the results of the data analysis, offering quantitative insights into the factors influencing IoT adoption. Section "Discussion" interprets the findings in the context of the existing literature, discussing the implications for theory and practice. Finally, the section "Conclusions" summarizes the study's main contributions, outlines its limitations, and suggests directions for future research. This structure is designed to provide a clear and logical progression through the topics covered, facilitating a comprehensive understanding of the study's scope and conclusions.

Related Work

The review of existing literature on the topic of IoT has explored both practical and theoretical aspects, suggesting a correlation between IoT and other factors such as self-efficacy, technology utilization, motivation, security, privacy, training, and more [13-19], indicating a relationship between the IoT and other elements including self-reliance, technological use, inspiration, safety, privacy, schooling, and beyond [13-19]. Likewise, research has examined the impact of IoT in conjunction with TAM and other external factors, and IoT with the help of TAM and external factors [20,21].

As IoT applications become more intricate, they can significantly impact learning. The difficulty is the IoT technology's quick development, which requires diverse skills ranging from developing IoT applications to incorporating devices into management systems that analyze device-generated data [14,22]. Previous research has highlighted the significance of IoT in addressing challenges students encounter when using modern IoT apps and gadgets. Potential solutions include focusing on computational thinking education, assisting students in solving challenges, and providing clear instructions and training to facilitate the integration of new students with IoT devices and encourage training in it.

A different approach was taken in another study, where a workshop was offered to address the importance of IoT. The results showed that students found the workshop highly satisfactory for learning about IoT, improving problem-solving skills, and enhancing problem-solving capabilities, while also finding it enjoyable [13-15].

Several outside variables, including drive, contentment, ease of use, effectiveness, involvement, and interest, have been investigated to explore the relationship between IoT and students' attitudes. The most significant factors are motivation and enjoyment regarding IoT technology, which are essential for its acceptance among students. While satisfaction and performance indirectly affect the application of IoT, greater contentment levels are not proportional to higher interactions [18,19]. Recent studies have used qualitative and quantitative methods to examine the uptake of IoT and have identified favorable attitudes, ease of use, contentment, affordability, basic knowledge, security, and privacy as crucial factors that influence its adoption. Apart from motivation and enjoyment, training and experience are also crucial in uptaking IoT devices, and educational training workshops can significantly improve the learning curve, digital learning, real-life applications, and problem-solving skills [16,17,23].

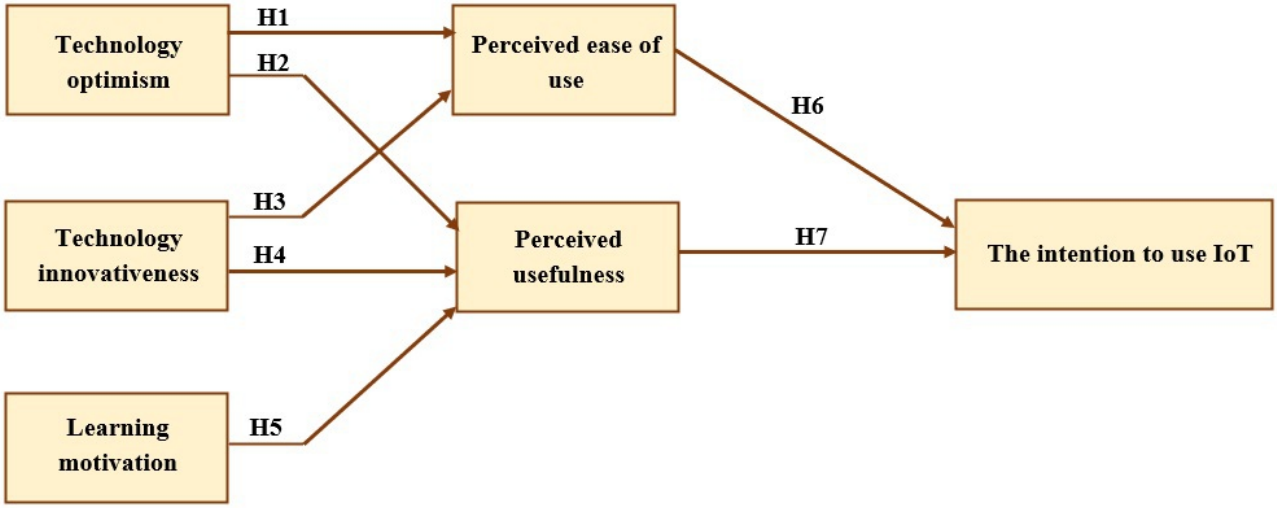
Previous research has explored the various benefits of implementing IoT technologies in educational settings, including the ability for teachers to gain insight into students' performance and knowledge levels, as well as the potential for improved teaching quality. As a result, the use of IoT has the potential to greatly impact learning environments, leading to a more

advanced educational landscape. This could also lead to changes in how we interact and collaborate as a society, as IoT technology continues to connect us in new ways [24,25]. While previous studies have investigated students' attitudes toward IoT from different angles [14,22,26], there has been limited research on the importance of IoT technology from the perspective of teachers' knowledge of the significance of IoT innovation from the educational viewpoint of instructors and how it impacts pedagogy in the classroom. Therefore, this study aims to address this gap by developing a model that combines both students' and teachers' attitudes to examine the effectiveness and efficiency of IoT in educational environments. By creating a framework that incorporates the viewpoints of learners and educators, this research intends to close this imbalance by analyzing the usefulness and efficacy of IoT in classrooms.

Developing Hypotheses and Theoretical Framework

The study paradigm shown in Figure 2 describes how the inspiration for learning, technological exuberance, and technological advancements influence learners' perceptions of the PU and ease of use (PE) of IoT devices. Such concepts have not yet been looked into in relation to IoT devices and applicability. Although earlier research has examined how these factors may affect IoT adoption plans [27], their individual effects on students' attitudes toward this technology have not been analyzed before.

Figure 2. Research model. IoT: Internet of Things.



LMT is a crucial factor impacting students' behavior in conventional and digital educational settings. It is an outside variable of the suggested approach and can be influenced by multiple factors, including the learning environment, expectations, and social values. Research has shown that LMT significantly impacts students' academic achievements [28]. Students must therefore be driven to acquire knowledge in order to achieve the suggested intended goals [29,30]. Highly motivated learners can achieve their learning goals spontaneously and willingly. Social cognitivism emphasizes the significance of requirements, conducts, and ideologies in education.

Researchers are investigating the influence of TOP and innovation on students' adoption of IoT in the education sector. The study aims to understand students' attitudes and behavior toward technology by using 3 theoretical concepts: inventiveness, technological exuberance, and learning incentive. The researchers have combined the (TAM) constructs with the aforementioned variables to accomplish their study's objectives. Several researchers have used the TAM model to evaluate the effectiveness of IoT technology and its applications [31,32]. The model explains why people accept technology based on their attitudes and beliefs and how these beliefs impact their behavioral intention toward technology use [33,34]. It is an established concept that supports embracing technology in various settings and strongly links it to technological progress [35-37]. The TAM model can forecast variables like technology elation and inventiveness that influence how well a certain technology is received [38,39]. TOP refers to users' favorable perception of technology, while technology innovativeness (TIN) refers to users' willingness to adopt technology early and lead to its use.

The PE and PU constructs of TAM are closely related to TOP and TIN. In the academic setting, students' perceptions of technology can be influenced by their peers and instructors. If the student's immediate academic circle has a positive view of a particular technology, the student will likely also develop a favorable opinion. Similarly, tech-savvy students tend to have a positive self-perception. Early on in the acceptance of technology, students are usually eager to pioneer using advanced technologies [40-44].

Furthermore, a student's optimism about technology is linked to their level of involvement in guiding its use. Similarly, technological optimism can significantly impact a student's attitude. IoT acceptance is greater for learners who are enthusiastic about experimenting with novel innovations. Technological innovators infrequently perceive new innovations as being challenging or outside their comprehension. Individuals are more inclined to regret not having the freedom to play around with novel technologies [45]. As a result, a number of hypotheses are put forth:

- H1: TOP affects positively the PE.
- H2: TOP affects positively the PU.
- H3: TIN affects positively the PE.
- H4: TIN affects positively the PU.
- H5: LMT affects positively the PU.
- H6: PE affects positively the intention to use Internet of Things (INT).
- H7: PU affects positively the INT.

By evaluating the association between TAM and its connected factors, this research seeks to add to a collection of current work. It has been said that it is important to measure these variables and analyze their relationship to TAM to assess the efficiency of IoT in classrooms. Past research and literature have mostly focused on preservice teachers, and there is a need to conduct more studies involving in-service teachers to enhance the practicality of the TAM model [46,47]. Thus, this study investigates the correlation between teachers' levels of TAM and their attitudes toward IoT acceptance when working with teachers from different fields.

Methods

Data Collection

The data collection for this study was carried out between January 20 and March 20, 2023, throughout the academic year 2023 - 2024's winter semester, at educational institutions in the United Arab Emirates. The research team used web-based surveys to collect data, with 500 questionnaires randomly distributed. Of these, 463 surveys were answered, resulting in a response rate of 93%. Some questionnaires were rejected due to missing values. Since a few surveys had no responses, they were discarded, leaving 769 usable questionnaires, which is considered an appropriate acceptable sample size according to Krejcie and Morgan [48]. Although the sample size exceeded the minimum requirements, the research team used SEM SmartPLS (version 3.2.7; SmartPLS GmbH) and SPSS Statistics (version 23; IBM Corp) to evaluate the hypotheses and confirm the relationship between variables. It is important to note that the hypotheses were based on previous theories related to IoT, which formed the foundation of this study.

Ethical Considerations

Ethical considerations were meticulously adhered to throughout the research process. All procedures involving human participants were approved by the institutional review board of the host universities in the United Arab Emirates (#RAREC00065), ensuring compliance with ethical standards. Informed consent was obtained from all participants before data collection, and they were informed of their rights to withdraw from the study at any time. Privacy and confidentiality were strictly maintained, with all data being anonymized and securely stored to prevent unauthorized access. Participants were not compensated for their participation, as the study involved minimal risk and was conducted as part of educational activities within the institutions involved.

Student's Personal Information

The demographic data of the respondents (N=463) are presented in Table 1. The data showed that 70% (n=325) of participants were female and 30% (n=138) were male. In terms of age, 42% (n=193) of participants were between 18 and 29 years, 35% (n=163) were between 30 and 39 years, 21% (n=99) were between 40 and 49 years, and 2% (n=8) were between 50 and 59 years. Regarding education, 80% (n=372) of participants held a bachelor's degree, 14% (n=62) held a master's degree, and 6% (n=29) held a doctoral degree. To obtain participants' willingness to participate, the research team used a purposive sampling approach. The participants came from various universities, academic levels, and programs relevant to this research. SPSS Statistics was used to analyze the demographic data.

Table . Demographic data of the respondents (N=463).

Category	Value, n (%)
Sex	
Female	325 (70)
Male	138 (30)
Age (years)	
Between 18 and 29	193 (42)
Between 30 and 39	163 (35)
Between 40 and 49	99 (21)
Between 50 and 59	8 (2)
Educational qualification	
Bachelor's degree	372 (80)
Master's degree	62 (14)
Doctorate	29 (6)

Study Instrument

This research proposed a survey instrument for validating the hypotheses. To assess the 6 constructs of the questionnaire, an additional 18 questions were included in the survey. The origins and histories of these constructs are presented in Table 2. To make the research more relevant, the researchers modified the questions from previous studies.

Table . Measurement items.

Constructs	Items	Definition	Instrument	Sources
TIN[a]	• TIN1 • TIN2 • TIN3	The term "technology innovativeness" describes a user's perception that they are at the forefront of technology use. Users who are pioneers in adopting new technologies typically do not view them as complicated or difficult to comprehend. Such users may feel a sense of regret if they miss the chance to experiment with new technologies.	• I accept IoT[b] technology to be used in my daily classes. • I am the only one. There is only me prepared to use IoT technology among my fellow students. • I am ready to use. I am prepared to use and experiment with the latest information technologies.	[45]
TOP[c]	• TOP1 • TOP2 • TOP3	A person's readiness to use technology is known as technological optimism.	• I am ready to test. Prepared to take the test IoT technology. • To complete my assignments to finish my homework, I will be using IoT. • I will learn more with the help of my preparedness to use IoT.	[2,49]
LMT[d]	• LMT1 • LMT2 • LMT3	The concept of learning motivation is used to measure the behavioral intention to use technology. Motivation learning is composed of 4 key components, which are attention, relevance, confidence, and satisfaction. These components have been identified in previous studies [50,51].	• I can improve my focus by using IoT for my daily classes. • Using IoT makes me feel more confident. • Using IoT for study purposes satisfies me.	[50,51]
PE[e]	• PE1 • PE2 • PE3	The TAM[f] was introduced by Davis [52] as a means of assessing the effectiveness and acceptance of technology. The model includes the concept of PE, which refers to the user's perception of how effortless it is to use the technology.	• IoT technology being simple will polish my skills. • I can improve my learning achievements by using IoT technology. • IoT is simple and easy to use.	[52]
PU[g]	• PU1 • PU2 • PU3	Usefulness refers to PU that the users of technology may see.	• IoT technology will hugely benefit me. • IoT will make my abilities and skills better. • Using IoT for my daily classes is beneficial.	[52]
INT[h]	• INT1 • INT2 • INT3	An individual's view of what others think about a certain behavior is known as a behavioral intention to use.	• IoT will be my go-to for daily tasks. • In the future, I will be using IoT. • IoT technology will be my recommendation to every student.	[53]

[a]TIN: technology innovativeness.

[b]IoT: Internet of Things.

[c]TOP: technology optimism.

[d]LMT: learning motivation.

[e]PE: perceived ease of use.

[f]TAM: technology acceptance model.

[g]PU: perceived usefulness.

[h]INT: intention to use Internet of Things.

Survey Structure

The questionnaire survey given to the students has 3 sections. Three components make up the survey that is provided to learners: (1) personal data is the focus of the first section, (2) the general question related to the "Intention to Use IoT" is the second section, and (3) 15 items that deal with TIN, TOP, LMT, PE, and PU is present in the third section.

A 5-point Likert scale was used to assess the 18 items, with response options ranging from 1=strongly disagree to 2=disagree, 3=neutral, 4=agree, and 5=strongly agree.

Results

Data Analysis

In this study, the gathered data were analyzed using SmartPLS (version 3.2.7) software through the partial least squares–structural equation modeling (PLS-SEM) technique [54-56]. The evaluation consisted of 2 stages: the measurement model and the structural model [57,58]. PLS-SEM was selected for this research after considering several factors.

The selection of PLS-SEM for this study was based on several reasons. First, PLS-SEM is preferred when the study aims to build on an existing theory [59]. Second, PLS-SEM is effective in handling complex models in exploratory research. Third, PLS-SEM analyzes the entire model as a single entity rather than dividing it into components [60]. Finally, PLS-SEM allows for the simultaneous analysis of structural and measurement models, leading to more accurate results [61].

Convergent Validity

Hair et al [57] suggested that to assess the measurement model, it is important to examine the construct reliability (including Cronbach ɑ and composite reliability) and validity (including convergent and discriminant validity). Table 3 shows that the Cronbach ɑ readings are greater than the suggested criterion of 0.7, and vary from 0.797 to 0.858 [62], indicating good construct reliability. Similarly, the proposed criterion is also greater than the composite reliability numbers, which vary from 0.735 to 0.858 [63]. Factor loading and average-variance extracted (AVE) analysis are required to evaluate convergent validity [57]. The factor loading values in Table 3 are higher than the recommended threshold of 0.7, and the AVE readings are greater than the suggested limit of 0.5, spanning from 0.556 to 0.712. These results suggest that there is convergent validity.

Table . Convergent validity results which assures acceptable values (factor loading, CA[a], CR[b], Dijkstra-Henseler's ρ ≥0.70, and AVE[c] >0.5).

Constructs and items	Factor loading	CA	CR	AVE
TIN[d]		0.856	0.824	0.608
TIN1	0.723			
TIN2	0.873			
TIN3	0.858			
TOP[e]		0.842	0.858	0.701
TOP1	0.804			
TOP2	0.816			
TOP3	0.801			
LMT[f]		0.815	0.851	0.712
LMT1	0.765			
LMT2	0.844			
LMT3	0.758			
PE[g]		0.797	0.735	0.662
PE1	0.858			
PE2	0.825			
PE3	0.758			
PU[h]		0.858	0.853	0.556
PU1	0.801			
PU2	0.829			
PU3	0.732			
INT[i]		0.825	0.843	0.612
INT1	0.812			
INT2	0.721			
INT3	0.749			

[a]CA: Cronbach ɑ.

[b]CR: composite reliability.

[c]AVE: average-variance extracted.

[d]TIN: technology innovativeness.

[e]TOP: technology optimism.

[f]LMT: learning motivation.

[g]PE: perceived ease of use.

[h]PU: perceived usefulness.

[i]INT: intention to use Internet of Things.

Discriminant Validity

To assess discriminant validity, 2 criteria were recommended: the Heterotrait-Monotrait ratio and the Fornell-Larcker criterion [57]. According to Table 4, the Fornell-Larcker criterion is met, as the AVE and its square root for each construct exceed its correlation with other constructs [64]. Table 5 displays the Heterotrait-Monotrait ratio results, which demonstrate that each construct's value is below the threshold value of 0.85 [65]. This suggests that discriminant validity exists, and the measurement model's reliability and validity were confirmed without any issues. As a result, the collected data can be used for analyzing the structural model.

Table . Fornell-Larcker scale.

	TIN[a]	TOP[b]	LMT[c]	PE[d]	PU[e]	INT[f]
TIN	0.864[g]	—[h]	—	—	—	—
TOP	0.675	0.883[g]	—	—	—	—
LMT	0.182	0.263	0.723[g]	—	—	—
PE	0.664	0.245	0.236	0.861[g]	—	—
PU	0.664	0.283	0.373	0.313	0.812[g]	—
INT	0.540	0.573	0.275	0.407	0.286	0.890[g]

[a]TIN: technology innovativeness.

[b]TOP: technology optimism.

[c]LMT: learning motivation.

[d]PE: perceived ease of use.

[e]PU: perceived usefulness.

[f]INT: intention to use Internet of Things.

[g]These values represent the square root of the average variance extracted for each construct, according to the Fornell-Larcker criterion. They are placed diagonally to demonstrate discriminant validity. A construct should share more variance with its indicators than with other constructs (off-diagonal correlations).

[h]Not applicable.

Table . Heterotrait-Monotrait ratio.

	TIN[a]	TOP[b]	LMT[c]	PE[d]	PU[e]	INT[f]
TIN	—[g]	0.355	0.473	0.113	0.741	0.336
TOP	0.355	—	0.406	0.512	0.579	0.512
LMT	0.473	0.406	—	0.702	0.559	0.021
PE	0.113	0.512	0.702	—	0.328	0.363
PU	0.741	0.579	0.559	0.328	—	0.486
INT	0.336	0.512	0.021	0.363	0.486	—

[a]TIN: technology innovativeness.

[b]TOP: technology optimism.

[c]LMT: learning motivation.

[d]PE: perceived ease of use.

[e]PU: perceived usefulness.

[f]INT: intention to use Internet of Things.

[g]Not applicable.

Hypotheses Testing Using PLS-SEM

To assess whether the theoretical constructs of the structural model are interconnected, the study used Smart PLS with maximum likelihood estimation to create a structural equation model [66]. The proposed hypotheses were then analyzed using this model. The results indicated a high level of predictive power for the model, with 86.3% of the variance in INT being accounted for, as shown in Figure 3 and Table 6.

Table 7 provides information on the β values, *t* values, and *P* values of all developed hypotheses based on the findings produced using the PLS-SEM technique. The researchers have confirmed each hypothesis. The empirical data supported H1, H2, H3, H4, H5, H6, and H7 following the data analysis hypotheses. This study demonstrates that TOP and TIN have a noteworthy influence on PE with respective regression coefficients (β=.412, *P*=.04), and (β=.608, *P*=.002), supporting H1 and H3. Moreover, the results indicate that PU is significantly impacted by TOP (β=.381, *P*=.04), TIN (β=.557, *P*=.003), and LMT (β=.752, *P*<.001), supporting hypotheses H2, H4, and H5, respectively. Finally, the study reveals that the relationship between PE and PU significantly affects INT with respective regression coefficients (β=.619, *P*<.001) and (β=.598, *P*<.001), supporting H6 and H7.

Figure 3. The path coefficients of the research model. IoT: Internet of Things. *P<.05; **P<.01; ***P<.001.

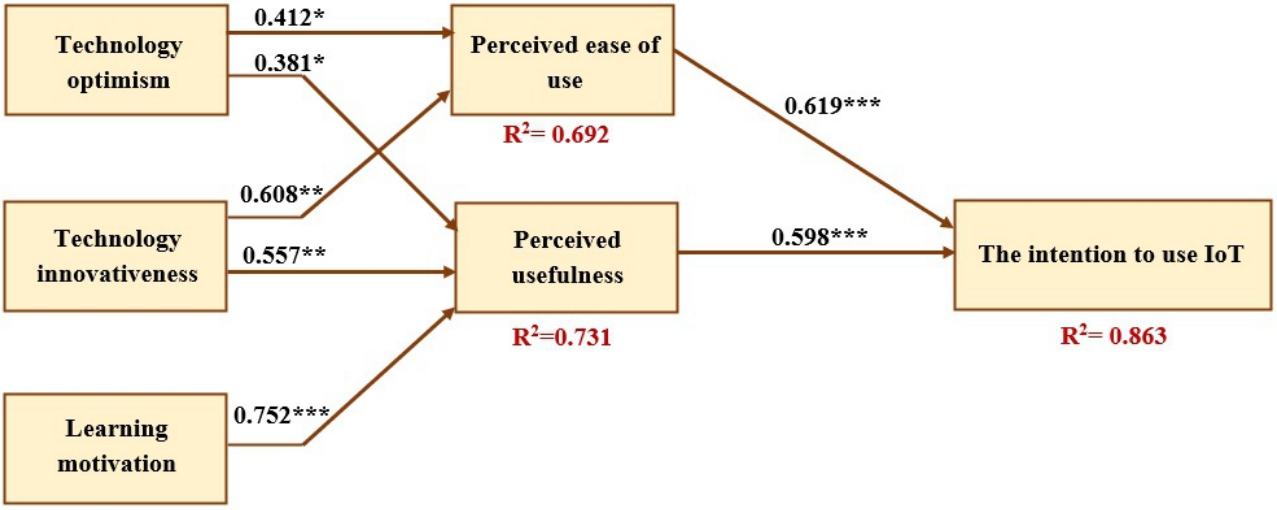


Table . The R^2 values for the endogenous latent variables.

Construct	R^2	Results
PE[a]	0.692	High
PU[b]	0.731	High
INT[c]	0.863	High

[a]PE: perceived ease of use.

[b]PU: perceived usefulness.

[c]INT: intention to use Internet of Things.

Table . Results of hypotheses testing for the research model at significance levels of .01 and .05.

H	Relationship	Path	*t* test (*df*)	*P* value	Direction	Decision
H1	TOP[a] -> PE[b]	0.412	5.552 (461)	.048	Positive	Supported
H2	TOP -> PU[c]	0.381	4.843 (461)	.04	Positive	Supported
H3	TIN[d] -> PE	0.608	10.247 (461)	.002	Positive	Supported
H4	TIN -> PU	0.557	9.358 (461)	.003	Positive	Supported
H5	LMT[e] -> PU	0.752	14.450 (461)	<.001	Positive	Supported
H6	PE -> INT[f]	0.619	16.753 (461)	<.001	Positive	Supported
H7	PU -> INT	0.598	14.195 (461)	<.001	Positive	Supported

[a]TOP: technology optimism.

[b]PE: perceived ease of use.

[c]PU: perceived usefulness.

[d]TIN: technology innovativeness.

[e]LMT: learning motivation.

[f]INT: intention to use Internet of Things.

ANN Results

The predictors identified during the PLS-SEM analysis are further investigated through ANN analysis using SPSS software. The PLS-SEM analysis identified TIN, TOP, LMT, PE, and PU as critical factors; therefore, ANN analysis also considers these 3 factors only. The structure of the ANN model is based on behavioral intention as an output neuron and TIN, TOP, LMT, PE, and PU as input neurons (Figures 4-6). ANN model supported deep learning in all the output neuron modes through its 2-hidden layer deep structure [67,68]. The researcher applied the activation function of the sigmoid function to hidden neurons, as well as output neurons, keeping the values of input and output neurons between [0, 1]; this allowed the researcher to obtain better performance from the research model [69,70]. He also applied the 10-fold cross-validation method to training

and testing data in the ratio of 80:20 to ensure that there is no overfitting in the ANN model [71]. The researcher evaluated the root mean square of error (RMSE) to test the neural network model for accuracy. The training data showed an RMSE value of 0.1388 for the ANN model while the testing data showed an RMSE value of 0.1439. The training and testing data showed only slight variance in the RMSE values and SD values (ie, 0.0043 and 0.0096). Hence, we can infer that using the ANN model enhances the accuracy of the research model.

Figure 4. ANN model for predicting PE. ANN: artificial neural network; PE: perceived ease of use; TIN: technology innovativeness; TOP: technology optimism.

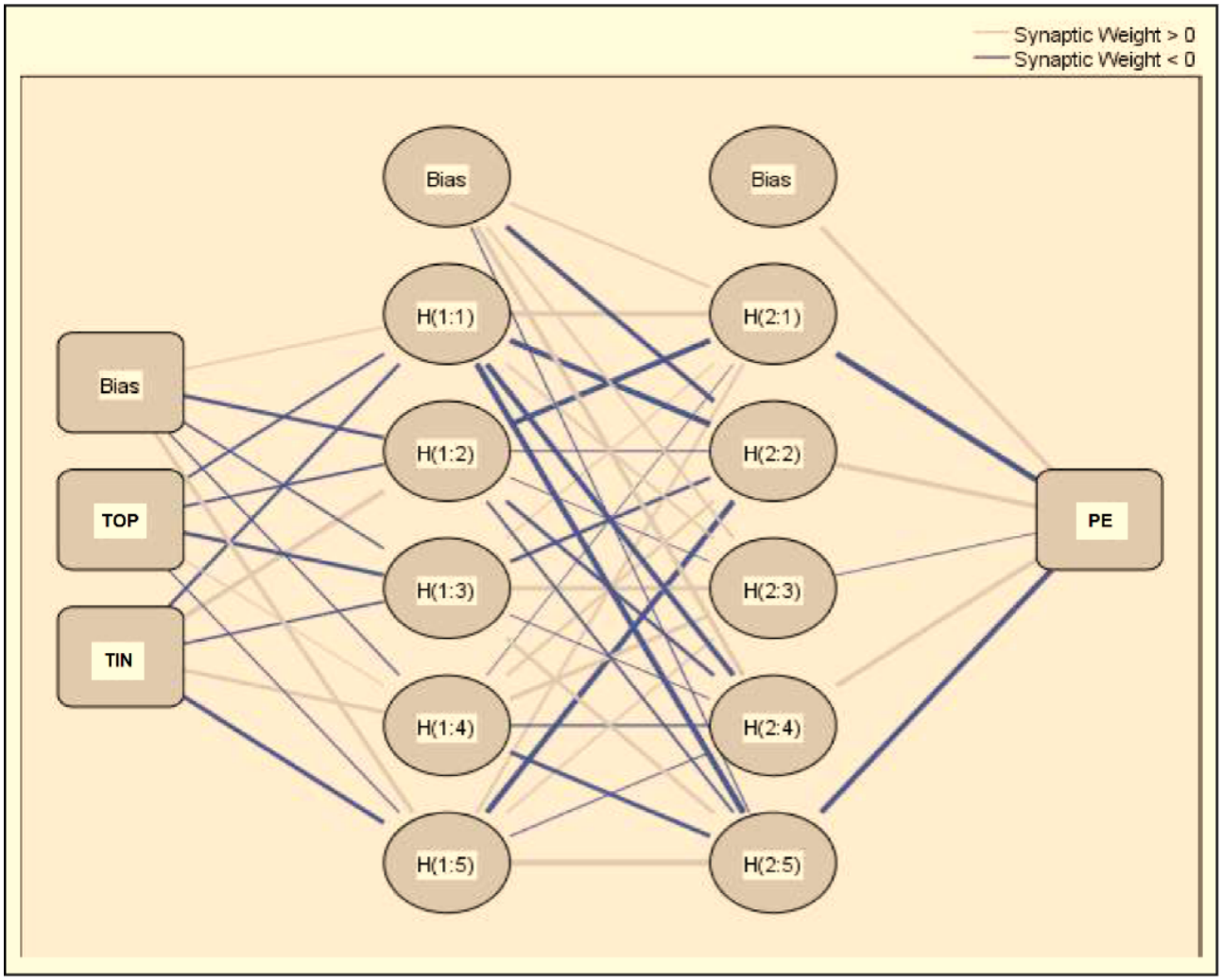


Figure 5. ANN model for predicting PU. ANN: artificial neural network; LMT: learning motivation; PU: perceived usefulness; TIN: technology innovativeness; TOP: technology optimism.

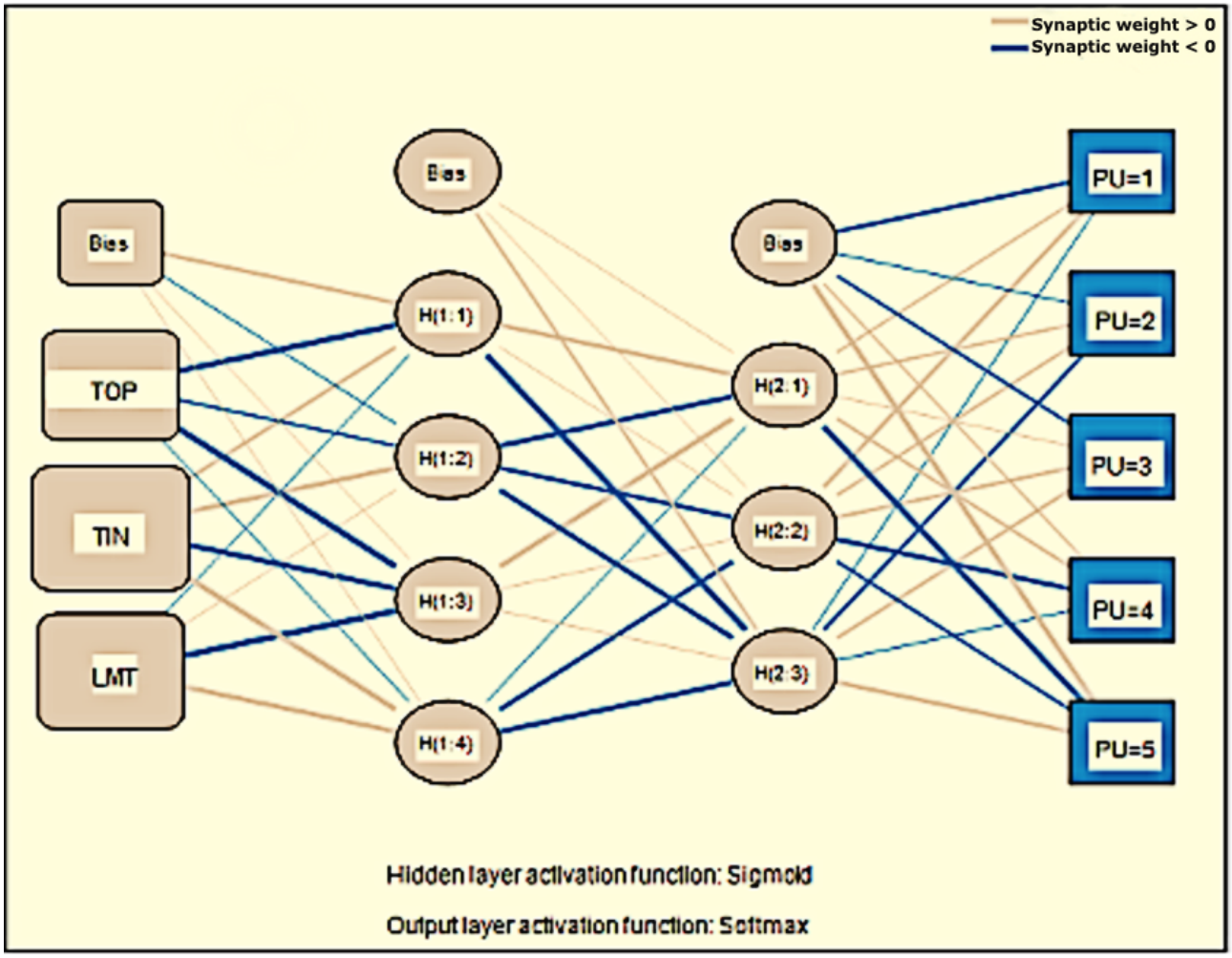
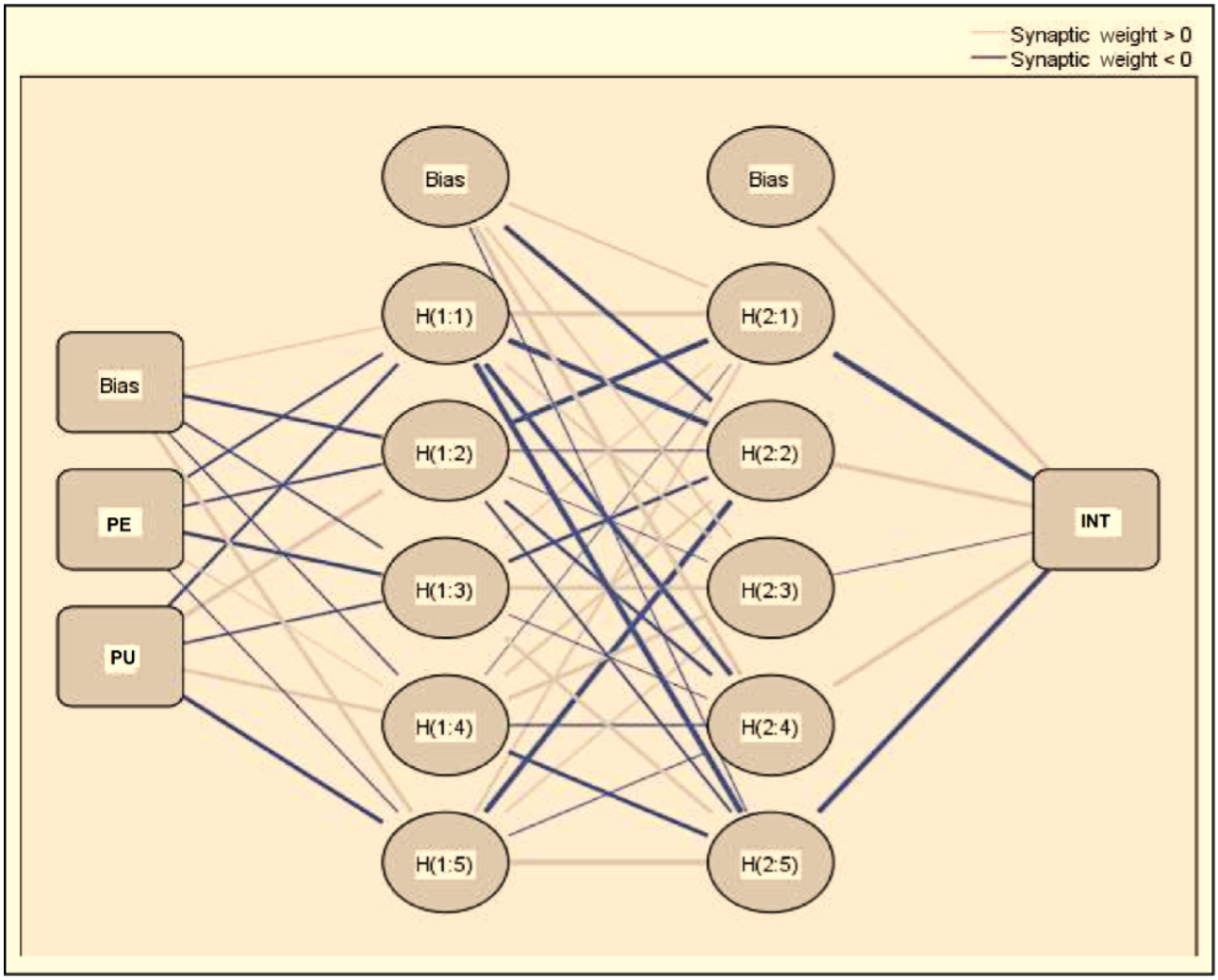


Figure 6. ANN model for predicting INT. ANN: artificial neural network; PE: perceived ease of use; PU: perceived usefulness; INT: intention to use Internet of Things.



Sensitivity Analysis

The researcher estimated the value of normalized importance by comparing each predictor's average value with the maximum mean value depicted among all predictors stated in percentage form. The values of normalized importance and mean importance computed for the predictors involved in ANN modeling are recorded in Table 8. This table also depicts the outcomes of the sensitivity analysis which identifies the predictor of PU to have the most significant impact on behavioral intention; the second most significant impact was imposed by TOP while the least impact was imposed by PE. ANN application was additionally assessed for its accuracy and performance by computing the goodness-of-fit for authentication and validation of the application. The ANN application uses goodness-of-fit just as PLS-SEM analysis uses R^2 [72,73]. The predictive powers of both the applications are compared where ANN analysis outperforms the other one with a predictive power of (R^2=89.7%) against the PLS-SEM predictive power of (R^2=86.3%). Hence, ANN analysis can explain endogenous constructs more effectively than PLS-SEM. Moreover, ANN analysis is based on deep learning and has the potential to better identify nonlinear associations among constructs which results in variances in the predictive powers of the ANN and PLS-SEM methods.

Table . Independent variable importance.

	Importance	Normalized importance (%)
TIN[a]	0.335	83.1
TOP[b]	0.463	94.8
LMT[c]	0.361	77.5
PE[d]	0.118	19.2
PU[e]	0.539	100

[a]TIN: technology innovativeness.

[b]TOP: technology optimism.

[c]LMT: learning motivation.

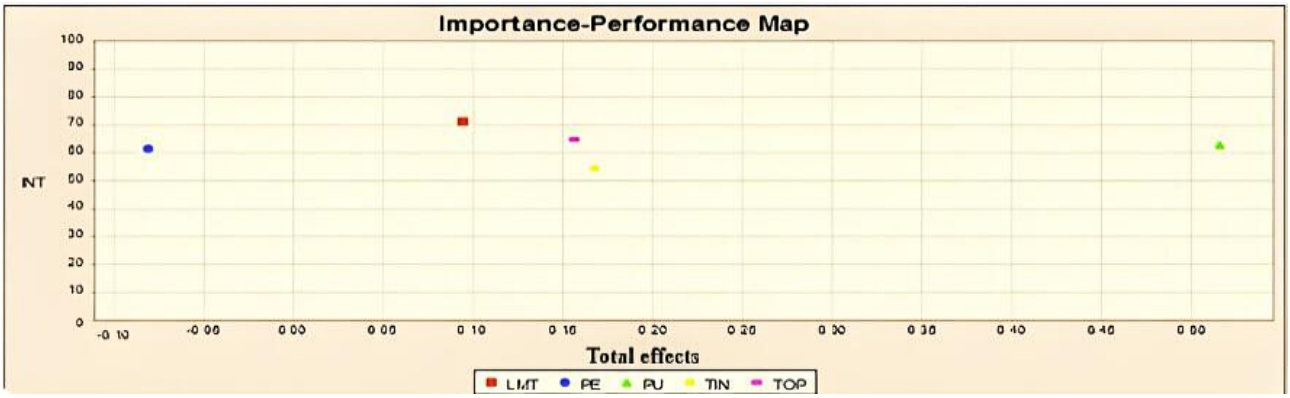
[d]PE: perceived ease of use.

[e]PU: perceived usefulness.

Importance-Performance Map Analysis

This study has used importance-performance map analysis (IPMA) as an advanced approach in PLS-SEM, which used behavioral intention as the main variable. As suggested by Ringle and Sarstedt [74], a better interpretation of the results of PLS-SEM is possible by using IPMA. A substitute way to only test the path coefficients (ie, importance measure), the average value of the latent constructs, and their indicators (ie, performance measure) are also included in the IPMA [74]. According to IPMA, the total effects reflect the predecessor factors' importance in developing the target factor (ie, behavioral intention), while the average of latent constructs' values is a reflection of their performance. The IPMA findings are reported in Figure 7. The estimation of the importance and performance of the 5 factors (ie, TIN, TOP, LMT, PE, and PU) has been shown in this table. According to the findings, the PU has been reported to have the largest values in terms of both importance and performance measures. Furthermore, it can be clearly seen that TIN has the second largest values in terms of both importance and performance measures. The third largest value was reported in the case of TOP in terms of the importance measure; however, it has the smallest value on the performance measure. Relatively, the opposite scenario was reported in the case of PE, as it had the lowest value on the importance measure.

Figure 7. Importance-performance map analysis results. INT: intention to use Internet of Things; LMT: learning motivation; PE: perceived ease of use; PU: perceived usefulness; TIN: technology innovativeness; TOP: technology optimism.



Discussion

Principal Findings

The main findings of this study indicate that the intention to adopt IoT technologies is positively influenced by both intrinsic motivational factors and TAM constructs. Specifically, TOP, innovation, and LMT play crucial roles at the initial level of influencing IoT adoption. At a secondary level, the perceived ease of use and PU, core elements of the TAM, directly enhance the INT technologies. Further analysis using ANNs and IPMA highlighted PU as a particularly significant predictor of IoT use intentions.

Based on these results, it is evident that technology features are critical in shaping users' PE and PU toward IoT applications and tools. The positive influence of TOP, innovation, and LMT is reflected across all model variables, suggesting that well-designed technology features can significantly boost IoT acceptance. The TAM framework effectively provides a user-friendly experience that meets users' value expectations and fosters positive emotional responses, which in turn, positively impacts their INT. Additionally, the significant positive impact of technology features on TOP boosts users' trust in the system's quality, further influencing their willingness to engage with IoT systems. The social aspect of technology use, where familiarity influences adoption, also plays a pivotal role in the INT features. Technology innovation contributes

significantly to perceived ease and usefulness, enhancing trust and satisfaction during the user experience, thus fostering a conducive environment for IoT adoption.

The confirmation of hypotheses H1-H5 supports the proposed conceptual model and the proposed hypotheses [75-77]. Previous research has also produced results that are consistent with these findings. However, other studies have shown that the lack of sufficient security and privacy are major challenges that may hinder the deployment of IoT in education [63,64], and to reduce these obstacles, future efforts to implement IoT in education must consider these factors. Although IoT has not been widely adopted in resource-limited countries, scholars need to examine the factors affecting its adoption to enable effective deployment. Therefore, scholars need to investigate the use and adoption of these technologies in other domains.

Managerial Implications

The adoption of IoT in educational settings demands that administrators and educational leaders ensure their faculty is proficient in using IoT technologies effectively in the classroom. This includes competencies in handling relevant technology tools, understanding pedagogical integration, and applying these technologies within various teaching scenarios. To support this, institutions should offer targeted professional development that focuses on both the technical and educational aspects of IoT. Additionally, management should consider the infrastructure upgrades necessary to support IoT technologies, such as improved wireless networks and enhanced security measures to protect student data [78,79].

Practical Implications

The practical applications of this study highlight the necessity for educational curricula to evolve alongside technological advancements. Institutions offering programs in computer science and engineering should integrate IoT courses to prepare students for the demands of the workforce, which increasingly relies on IoT technologies. Moreover, schools should align their IoT strategies with real-world applications, providing students with hands-on opportunities to work with IoT in context. This could include partnerships with IoT companies or practical projects that allow students to solve real problems using IoT solutions [80,81].

Theoretical Implications

This study contributes to the academic understanding of technology adoption by confirming the significant role of intrinsic motivational factors and TAM constructs in the adoption of IoT technologies. By highlighting the dual influence of personal motivation and perceived technological attributes, this research extends existing models of technology acceptance. Furthermore, the use of advanced analytical methods such as ANN and IPMA provides a deeper insight into the nonlinear relationships among the constructs, offering a nuanced perspective that can inform future research in technology adoption theories. This could encourage scholars to explore how different educational contexts or cultural backgrounds influence the adoption and effective use of emerging technologies like IoT [82,83].

Conclusions

IoT technology has fundamentally altered the tech and business industries, laying the groundwork for the creation of intelligent societies and advancing social and economic development. The IoT has advanced quickly and significantly. The research team developed the PE and PU frameworks for Arab customers using path estimation and modeling of structural equations depending on their responses to analyze individuals' acceptance of IoT. In addition to additional factors including technology exuberance, advancements in technology, and academic drive, the research additionally looked at the effects of TAM components on these factors. These factors and TAM components were found to be directly related to the study. According to the research, the incentive to learn, technological advancement, and technological positiveness all had a substantial influence on PE and PU. While earlier research has demonstrated that TAM components affect the motivation to use IoT, this research has investigated how other external factors, including social factors like LMT, can affect these constructs. The study used a conceptual model to examine users' attitudes toward IoT adoption, with 2 levels of analysis. The first level focused on social attitudes, particularly the impact of the incentive to learn as a standalone factor. The second stage looked at how personal traits influenced technological positivism and inventiveness, influencing customer needs and IoT interactions. Subsequent research could look at how individual characteristics affect incentives to learn and look at extra technological aspects that affect IoT uptake. Subsequent research could improve the assessment and give more insight into the value of IoT by including mediating factors between personal characteristics and technological attributes. The theoretical framework could also be used with cutting-edge technology like metaverse systems and artificial intelligence.

Data Availability

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

Authors' Contributions

K Alhumaid wrote the manuscript and conducted the formal analysis. SS validated the results and reviewed and edited the manuscript. MK developed the software and created the visualizations. K Ayoubi contributed to the investigation and validation. All authors reviewed and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Ali M, Bilal HSM, Razzaq MA, et al. IoTFLiP: IoT-based flipped learning platform for medical education. Digital Commun Netw 2017 Aug;3(3):188-194. [doi: 10.1016/j.dcan.2017.03.002]
2. Almaiah M, Alfaisal R, Salloum S, et al. Integrating teachers' TPACK levels and students' learning motivation, technology innovativeness, and optimism in an IoT acceptance model. Electronics (Basel) 2022;11(19):3197. [doi: 10.3390/electronics11193197]
3. Al-Adwan AS, Al-Debei MM. The determinants of Gen Z's metaverse adoption decisions in higher education: integrating UTAUT2 with personal innovativeness in IT. Educ Inf Technol 2024 Apr;29(6):7413-7445. [doi: 10.1007/s10639-023-12080-1]
4. Banica L, Burtescu E, Enescu F. The impact of internet-of-things in higher education. Sci Bull Sci 2017;16(1):53-59 [FREE Full text]
5. Gul S, Asif MA, Ahmad S, et al. A survey on role of internet of things in education. Int J Comput Sci Netw Secur 2017;17(5):159-165 [FREE Full text]
6. Wakefield KJ. How the Internet of Things is transforming manufacturing. Forbes. 2014 Jul 1. URL: https://www.forbes.com/sites/ptc/2014/07/01/how-the-internet-of-things-is-transforming-manufacturing/ [accessed 2025-03-26]
7. Aldowah H, Ul Rehman S, Ghazal S, Naufal Umar I. Internet of Things in higher education: a study on future learning. J Phys: Conf Ser 2017 Sep 21;892(1):012017. [doi: 10.1088/1742-6596/892/1/012017]
8. Ning H, Hu S. Technology classification, industry, and education for future Internet of Things. Int J Commun 2012 Sep;25(9):1230-1241. [doi: 10.1002/dac.2373]
9. Kurdi BA, Alshurideh A, Nuseir M, Aburayya A, Salloum SA. The effects of subjective norm on the intention to use social media networks: an exploratory study using PLS-SEM and machine learning approach. In: Advanced Machine Learning Technologies and Applications (AMLTA 2021) 2021:581-592. [doi: 10.1007/978-3-030-69717-4_55]
10. Almarzouqi A, Aburayya A, Salloum SA. Determinants predicting the electronic medical record adoption in healthcare: a SEM-artificial neural network approach. PLoS One 2022;17(8):e0272735. [doi: 10.1371/journal.pone.0272735] [Medline: 35972979]
11. Almarzouqi A, Aburayya A, Salloum SA. Determinants of intention to use medical smartwatch-based dual-stage SEM-ANN analysis. Inform Med Unlocked 2022;28:100859. [doi: 10.1016/j.imu.2022.100859]
12. Alhumaid K, Alnazzawi N, Akour I, Khasoneh OA, Alfaisal R, Salloum S. An integrated model for the usage and acceptance of stickers in WhatsApp through SEM-ANN approach. Int J Data Netw Sci 2022;6(4):1261-1272. [doi: 10.5267/j.ijdns.2022.6.008]
13. Malhotra S, Kumar A, Dutta R. Effect of integrating IoT courses at the freshman level on learning attitude and behaviour in the classroom. Educ Inf Technol 2021 May;26(3):2607-2621. [doi: 10.1007/s10639-020-10376-0]
14. Lai YH, Chen SY, Lai CF, Chang YC, Su YS. Study on enhancing AIoT computational thinking skills by plot image-based VR. Interact Learn Environ 2021 Apr 3;29(3):482-495. [doi: 10.1080/10494820.2019.1580750]
15. Mavroudi A, Divitini M, Gianni F, Mora S, Kvittem DR. Designing IoT applications in lower secondary schools. Presented at: 2018 IEEE Global Engineering Education Conference (EDUCON); Apr 17-20, 2018; Santa Cruz de Tenerife, Spain.
16. Yamamoto T, Horita T. Analysis of teacher training students' understanding on learning objectives in elementary school programming classes using IoT materials. Presented at: Society for Information Technology & Teacher Education International Conference; Apr 6-10, 2020; New Orleans, LA, United States.
17. Akbar MA, Rashid MM, Embong AH. Technology based learning system in Internet of Things (IoT) education. Presented at: 2018 7th International Conference on Computer and Communication Engineering (ICCCE); Sep 19-20, 2018; Kuala Lumpur. [doi: 10.1109/ICCCE.2018.8539334]
18. Lee S, Cho K. A study on the structural relationship between IoT usage and life satisfaction among university students. J Internet Things Converg 2021;7(2):55-63 [FREE Full text]
19. Glaroudis D, Iossifides A, Spyropoulou N, Zaharakis ID. Investigating secondary students' stance on IoT driven educational activities. Presented at: European Conference on Ambient Intelligence; Nov 12-14, 2018; Larnaca, Cyprus.
20. Yildiz Durak H. Modeling of relations between K-12 teachers' TPACK levels and their technology integration self-efficacy, technology literacy levels, attitudes toward technology and usage objectives of social networks. Interact Learn Environ 2021 Oct 3;29(7):1136-1162. [doi: 10.1080/10494820.2019.1619591]
21. Yang J, Wang Q, Wang J, Huang M, Ma Y. A study of K-12 teachers' TPACK on the technology acceptance of E-schoolbag. Interact Learn Environ 2021 Oct 3;29(7):1062-1075. [doi: 10.1080/10494820.2019.1627560]
22. Pastor-Vargas R, Tobarra L, Robles-Gómez A, Martin S, Hernández R, Cano J. A WoT platform for supporting full-cycle IoT solutions from edge to cloud infrastructures: a practical case. Sensors (Basel) 2020 Jul 5;20(13):3770. [doi: 10.3390/s20133770] [Medline: 32635632]
23. Akbari M, Rezvani A, Shahriari E, Zúñiga M, Pouladian H. Acceptance of 5 G technology: mediation role of trust and concentration. J Eng Technol Manag 2020 Jul;57:101585. [doi: 10.1016/j.jengtecman.2020.101585]
24. Seiler L, Kuhnel M, Honal A, Ifenthaler D. Mobile learning analytics: potenziale für lernen und lehren am beispiel hochschule. In: Handbuch Mobile Learning: Springer; 2018:585-608.

25. Majeed A, Ali M. How internet-of-things (iot) making the university campuses smart? QA higher education (QAHE) perspective. Presented at: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC); Jan 8-10, 2018; Las Vegas, NV, United States. [doi: 10.1109/CCWC.2018.8301774]
26. Singh UG, Msibi A. Students' perceptions towards IoTs adoption: a case study. J Phys: Conf Ser 2021 Jan 1;1714(1):012044. [doi: 10.1088/1742-6596/1714/1/012044]
27. Liu Y, Gan Y, Song Y, Liu J. What influences the perceived trust of a voice-enabled smart home system: an empirical study. Sensors (Basel) 2021 Mar 13;21(6):2037. [doi: 10.3390/s21062037] [Medline: 33805702]
28. Hampton NZ, Mason E. Learning disabilities, gender, sources of efficacy, self-efficacy beliefs, and academic achievement in high school students. J Sch Psychol 2003 Mar;41(2):101-112. [doi: 10.1016/S0022-4405(03)00028-1]
29. McCown RR, Driscoll M, Roop P. Facilitating student motivation. In: Educational Psychology: A Learning-Centered Approach to Classroom Practice: Allyn and Bacon; 1996:278-309.
30. Stipek D, Feiler R, Daniels D, Milburn S. Effects of different instructional approaches on young children's achievement and motivation. Child Dev 1995 Feb;66(1):209. [doi: 10.2307/1131201]
31. Sırakaya M, Sırakaya DA. Augmented reality in STEM education: a systematic review. Interact Learn Environ 2020:1-14 [FREE Full text]
32. Alsyouf A, Lutfi A, Alsubahi N, et al. The use of a technology acceptance model (TAM) to predict patients' usage of a personal health record system: the role of security, privacy, and usability. Int J Environ Res Public Health 2023 Jan 11;20(2):1347. [doi: 10.3390/ijerph20021347] [Medline: 36674105]
33. Esteban-Millat I, Martínez-López FJ, Pujol-Jover M, Gázquez-Abad JC, Alegret A. An extension of the technology acceptance model for online learning environments. Interact Learn Environ 2018 Oct 3;26(7):895-910. [doi: 10.1080/10494820.2017.1421560]
34. Huang HM, Liaw SS, Lai CM. Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems. Interact Learn Environ 2016 Jan 2;24(1):3-19. [doi: 10.1080/10494820.2013.817436]
35. Parasuraman A. Technology Readiness Index (TRI) a multiple-item scale to measure readiness to embrace new technologies. J Serv Res 2000;2(4):307-320. [doi: 10.1177/109467050024001]
36. Al-Adwan AS. Investigating the drivers and barriers to MOOCs adoption: the perspective of TAM. Educ Inf Technol 2020 Nov;25(6):5771-5795. [doi: 10.1007/s10639-020-10250-z]
37. Alsyouf A, Lutfi A, Al-Bsheish M, et al. Exposure detection applications acceptance: the case of COVID-19. Int J Environ Res Public Health 2022 Jun 14;19(12):7307. [doi: 10.3390/ijerph19127307] [Medline: 35742560]
38. Chin WW. The partial least squares approach to structural equation modeling. In: Modern Methods for Business Research: Psychology Press; 1998, Vol. 295:295-336.
39. Alhumaid K, Rizk RY, Snášel V, Abdel-Kader RF. Predicting the intention to use audi and video teaching styles: an empirical study with PLS-SEM and machine learning models. Presented at: International Conference on Advanced Machine Learning Technologies and Applications; May 5-7, 2022; Cairo, Egypt. [doi: 10.1007/978-3-031-03918-8_23]
40. Ngafeeson MN, Sun J. E-Book acceptance among undergraduate students: a look at the moderating role of technology innovativeness. Int J Web-Based Learn Teach Technol 2015;10(3):36-51. [doi: 10.4018/ijwltt.2015070103]
41. Li CH, Wu JJ. The structural relationships between optimism and innovative behavior: understanding potential antecedents and mediating effects. Creat Res J 2011 Apr;23(2):119-128. [doi: 10.1080/10400419.2011.571184]
42. Ziyae B, Mobaraki MH, Saeediyoun M. The effects of psychological capital on organizational performance with an emphasis on the mediatory role of innovation in information technology: a case study of the branches of agriculture bank in Tehran. Innov Manag J 2015;3(4):77-100 [FREE Full text]
43. Issham I, Siti NA, Nizuwan A. Accessing innovativeness of distance learners toward their readiness in embracing technology. African J Bus Manag 2011;5(33):12768-12776. [doi: 10.5897/AJBM11.824]
44. Al-Adwan AS, Li N, Al-Adwan A, Abbasi GA, Albelbisi NA, Habibi A. Extending the Technology Acceptance Model (TAM) to predict university students' intentions to use metaverse-based learning platforms. Educ Inf Technol (Dordr) 2023 Apr 28:1-33. [doi: 10.1007/s10639-023-11816-3] [Medline: 37361794]
45. Karahanna E, Straub DW, Chervany NL. Information technology adoption across time: a cross-sectional comparison of pre-adoption and post-adoption beliefs. MIS Q 1999 Jun;23(2):183. [doi: 10.2307/249751]
46. Kabakci Yurdakul I, Odabasi HF, Kilicer K, Coklar AN, Birinci G, Kurt AA. The development, validity and reliability of TPACK-deep: a technological pedagogical content knowledge scale. Comput Educ 2012 Apr;58(3):964-977. [doi: 10.1016/j.compedu.2011.10.012]
47. Schmidt DA, Baran E, Thompson AD, Mishra P, Koehler MJ, Shin TS. Technological pedagogical content knowledge (TPACK) the development and validation of an assessment instrument for preservice teachers. J Res Technol Educ 2009;42(2):123-149. [doi: 10.1080/15391523.2009.10782544]
48. Krejcie RV, Morgan DW. Determining sample size for research activities. Educ Psychol Meas 1970 Sep;30(3):607-610. [doi: 10.1177/001316447003000308]
49. Chung N, Han H, Joun Y. Tourists' intention to visit a destination: the role of augmented reality (AR) application for a heritage site. Comput Human Behav 2015 Sep;50:588-599. [doi: 10.1016/j.chb.2015.02.068]

50. Huang HMM, Rauch U, Liaw SSS. Investigating learners' attitudes toward virtual reality learning environments: based on a constructivist approach. Comput Educ 2010 Nov;55(3):1171-1182. [doi: 10.1016/j.compedu.2010.05.014]
51. Huang HM, Liaw SS. An analysis of learners' intentions toward virtual reality learning based on constructivist and technology acceptance approaches. Int Rev Res Open Distrib Learn 2018;19(1):1. [doi: 10.19173/irrodl.v19i1.2503]
52. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
53. Fishbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research: Addison-Wesley; 1977.
54. Ringle CM, Wende S, Becker JM. SmartPLS 3. : SmartPLS; 2015.
55. Mouzaek E, Marzouqi A, Alaali N, Salloum S, Aburayya A, Suson O. An empirical investigation of the impact of service quality dimensions on guests satisfaction: a case study of Dubai hotels. J Contemp Issues Bus Gov 2021 Apr 2;27(3):1187. [doi: 10.47750/cibg.2021.27.03.160]
56. Salloum SA, Mhamdi C, Kurdi BA, Shaalan K. Factors affecting the adoption and meaningful use of social media: a structural equation modeling approach. Int J Inf Technol Lang Stud 2018;2(3):96-109.
57. Hair J, Hollingsworth CL, Randolph AB, Chong AYL. An updated and expanded assessment of PLS-SEM in information systems research. Ind Manag Data Syst 2017 Apr 10;117(3):442-458. [doi: 10.1108/IMDS-04-2016-0130]
58. Salloum SA, Al-Emran M, Habes M, Alghizzawi M, Ghani MA, Shaalan K. What impacts the acceptance of E-learning through social media? An empirical study. Recent Adv Technol Accept Model Theor 2021:419-431. [doi: 10.1007/978-3-030-64987-6_24]
59. MacKenzie SB, Podsakoff PM, Jarvis CB. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. J Appl Psychol 2005 Jul;90(4):710-730. [doi: 10.1037/0021-9010.90.4.710] [Medline: 16060788]
60. Goodhue DL, Lewis W, Thompson R. Does PLS have advantages for small sample size or non-normal data? MIS Q 2012;36(3):981. [doi: 10.2307/41703490]
61. Barclay D, Higgins C, Thompson R. The partial least squares (PLS) approach to casual modeling: personal computer adoption and use as an illustration. Technol Stud 1995 [FREE Full text]
62. Nunnally JC, Bernstein IH. Psychometric Theory: McGraw Hill; 1994.
63. Kline RB. Principles and Practice of Structural Equation Modeling: Guilford Publications; 2015.
64. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 1981 Feb;18(1):39-50. [doi: 10.1177/002224378101800104]
65. Henseler J, Ringle CM, Sarstedt M. A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 2015 Jan;43(1):115-135. [doi: 10.1007/s11747-014-0403-8]
66. Almaiah MA, Alfaisal R, Salloum SA, et al. Examining the impact of artificial intelligence and social and computer anxiety in E-learning settings: students' perceptions at the university level. Electronics (Basel) 2022;11(22):3662. [doi: 10.3390/electronics11223662]
67. Lee VH, Hew JJ, Leong LY, Tan GWH, Ooi KB. Wearable payment: a deep learning-based dual-stage SEM-ANN analysis. Expert Syst Appl 2020 Nov;157:113477. [doi: 10.1016/j.eswa.2020.113477]
68. Elareshi M, Habes M, Youssef E, Salloum SA, Alfaisal R, Ziani A. SEM-ANN-based approach to understanding students' academic-performance adoption of YouTube for learning during Covid. Heliyon 2022 Apr;8(4):e09236. [doi: 10.1016/j.heliyon.2022.e09236] [Medline: 35434400]
69. Liébana-Cabanillas F, Marinkovic V, de Luna IR, Kalinic Z. Predicting the determinants of mobile payment acceptance: a hybrid SEM-neural network approach. Technol Forecast Soc Change 2018 Apr;129:117-130. [doi: 10.1016/j.techfore.2017.12.015]
70. Aburayya A, Salloum SA, Alderbashi KY, et al. SEM-machine learning-based model for perusing the adoption of metaverse in higher education in UAE. Int J Data Netw Sci 2023;7(2):667-676. [doi: 10.5267/j.ijdns.2023.3.005]
71. Sharma SK, Sharma M. Examining the role of trust and quality dimensions in the actual usage of mobile banking services: an empirical investigation. Int J Inf Manage 2019 Feb;44:65-75. [doi: 10.1016/j.ijinfomgt.2018.09.013]
72. Leong LY, Hew TS, Ooi KB, Lee VH, Hew JJ. A hybrid SEM-neural network analysis of social media addiction. Expert Syst Appl 2019 Nov;133:296-316. [doi: 10.1016/j.eswa.2019.05.024]
73. Salloum SA, AlAhbabi NMN, Habes M, Aburayya A, Akour I. Predicting the intention to use social media sites: a Hybrid SEM - machine learning approach. Adv Intell Syst Comput 2021;1339:324-334. [doi: 10.1007/978-3-030-69717-4_32]
74. Ringle CM, Sarstedt M. Gain more insight from your PLS-SEM results: the importance-performance map analysis. Ind Manage Data Syst 2016. [doi: 10.1108/IMDS-10-2015-0449]
75. Zaidi MFA. The IoT readiness of SMEs in Malaysia: are they worthwhile for investigation? Presented at: International Conference on International Business, Marketing and Humanities (ICIBMAH 2017); Aug 26-27, 2017; Alor Setar, Malaysia.
76. Cadenas GA, Cantú EA, Lynn N, Spence T, Ruth A. A programmatic intervention to promote entrepreneurial self-efficacy, critical behavior, and technology readiness among underrepresented college students. J Vocat Behav 2020 Feb;116:103350. [doi: 10.1016/j.jvb.2019.103350]

77. Wang Y, Liu C, Tu YF. Factors affecting the adoption of AI-based applications in higher education. Educ Technol Soc 2021;24(3):116-129 [FREE Full text]
78. Chweya R, Ajibade SSM, Buba AK, et al. IoT and big data technologies: opportunities and challenges for higher learning. Int J Recent Technol Eng (IJRTE) 2020;9(2):909-913. [doi: 10.35940/ijrte.B3943.079220]
79. Rajamohan K, Rangasamy S, Manoj CN, Mary L, Sabu R, Jose SS. Internet of things (IoT) as a game changer to the education sector. In: Advanced IoT Technologies and Applications in the Industry 40 Digital Economy: CRC Press; 2024:251-274. [doi: 10.1201/9781003434269-14]
80. Abraham S. Using internet of things (IoT) as a platform to enhance interest in electrical and computer engineering. Presented at: 2016 ASEE Annual Conference & Exposition; Jun 26-29, 2016; New Orleans, Louisiana. [doi: 10.18260/p.27149]
81. Almodaires A. Technology-supported reflection: towards bridging the gap between theory and practice in teacher education. : University of Twente; 2009 URL: https://research.utwente.nl/en/publications/technology-supported-reflection-towards-bridging-the-gap-between--2 [accessed 2025-03-26]
82. Marikyan M, Papagiannidis P. Unified Theory of Acceptance and Use of Technology: a review. In: TheoryHub Book 2021.
83. Syed H, Malik AN. Comparative study of effect of culture on technology adoption in Pakistan and USA. Bus Manag Rev 2014;5(1):42 [FREE Full text]

Abbreviations

ANN: artificial neural network
AVE: average-variance extracted
INT: intention to use Internet of Things
IoT: Internet of Things
IPMA: importance-performance map analysis
LMT: learning motivation
PE: perceived ease of use
PLS-SEM: partial least squares–structural equation modeling
PU: perceived usefulness
RMSE: root mean square of error
SEM: structural equation modeling
TAM: technology acceptance model
TIN: technology innovativeness
TOP: technology optimism

Edited by A Kushniruk; submitted 14.03.24; peer-reviewed by A Lutfi, AS Al-Adwan, A Arbabisarjou, G Farid, R Alfaisal; revised version received 01.09.24; accepted 15.01.25; published 10.04.25.

Please cite as:
Alhumaid K, Ayoubi K, Khalifa M, Salloum S
Factors Determining Acceptance of Internet of Things in Medical Education: Mixed Methods Study
JMIR Hum Factors 2025;12:e58377
URL: https://humanfactors.jmir.org/2025/1/e58377
doi:10.2196/58377

The Impact of Trust and the Role of the Opt-Out Mechanism in Willingness to Share Health Data via Electronic Health Records in Germany: Telephone Survey Study

Felix Wilke, Prof Dr

Department of Social Work, University of Applied Sciences Jena, Carl-Zeiss-Promenade 2, Jena, Germany

Corresponding Author:
Felix Wilke, Prof Dr
Department of Social Work, University of Applied Sciences Jena, Carl-Zeiss-Promenade 2, Jena, Germany

Abstract

Background: Electronic health records (EHRs) offer a valuable resource for research and health care improvement. However, public acceptance of sharing personal health data is critical to the success of such initiatives. In Germany, automatic data sharing via EHRs will be implemented unless people opt out.

Objective: This study aims to assess the willingness of the German population to share health data via EHRs and to explore the role of trust in influencing these attitudes.

Methods: A computer-assisted telephone interview study was conducted in December 2023, with 1004 respondents aged 18 years and older, representative of the German population. Descriptive statistics and multivariate linear regression models were used to analyze the data.

Results: The survey shows that 43.4% (n=432) of respondents would be willing to share their health data via EHR, and a significant 34% (n=338) remain undecided. While the population is open to adoption of the EHR for personal health issues (n=483, 53% show interest in using it), the opt-out model for data sharing is viewed critically, with 44.7% (n=438) of respondents rejecting it. Socioeconomic status significantly influences the willingness to share data, with higher income, education, and digital literacy being associated with greater openness to data sharing. However, trust emerged as the most significant factor. Additionally, experiences with digital technologies increase the willingness to share personal health data.

Conclusions: The German population shows general openness toward EHRs and data sharing. Trust plays a critical role in promoting willingness to share health data. The findings highlight challenges in Germany's transition to an opt-out system.

(JMIR Hum Factors 2025;12:e65718) doi:10.2196/65718

KEYWORDS

data sharing; health; citizens; electronic health record; trust; digitalization; opt-out

Introduction

Legislative Context and Research Questions

Health data stored in electronic health records (EHRs) are considered an extremely valuable source for advancing medical science and health care. EHRs have the potential to provide high-quality longitudinal data from a broad spectrum of the population. It is therefore not surprising that many countries are implementing data-sharing options to make use of these data. In Germany, the EHR ("elektronische Patientenakte") was introduced on January 1, 2021, and is available to all legally insured persons through their health insurance company. It can be used to store and manage medical diagnoses and treatment data. The use of the EHR also allows control over who can view and access data. However, by 2024, only 1% - 2% of the eligible population have an EHR. To increase the prevalence and benefits of the EHR, the German Bundestag passed the Digital Act ("Digital-Gesetz") and the Health Data Use Act ("Gesundheitsdatennutzungsgesetz") on December 14, 2023. This means that starting in 2025, for all people with statutory health insurance in Germany, an EHR will be automatically created unless they opt out. The Health Data Use Act further allows for the use of pseudonymized health data for research purposes by public and private institutions by amending §303e SGB V. That means data stored in the EHR will be available for research purposes including commercial purposes unless people opt out. The data will be made available through the German Health Research Data Center ("Forschungsdatenzentrum"). The legislative changes mark a paradigm shift in the way health data are handled: the creation of electronic patient records and the sharing of health data will be automatic unless people actively opt out. The German federal minister of health, Karl Lauterbach, stated that the paradigm shift will be a "gamechanger" for the development of health [1]. The new legal framework marks a step toward digitization

that many experts have been calling for for a long time [2]. At the same time, the extensive and passive sharing of health data has been criticized by various stakeholders.

This paradigm shift raises some serious questions about the social implications. There are a number of challenges to the "data is gold" perspective [3]. Some social scientists criticize that digitization in health bears potential for "panoptic surveillance" [4]. Other critics [5] have highlighted data security issues that can harm citizens in many ways. Furthermore, the risks, but also the opportunities to benefit from digital health, are unevenly distributed across society. There is ample empirical evidence of a so-called "digital divide," meaning that people with lower socioeconomic status have lower chances of using digital (health) services and face higher risks [4,6,7]. The digitization of health therefore needs to be critically monitored. Citizens should be aware of the potential benefits as well as the potential dangers of sharing personal health data. This becomes even more urgent with the implementation of an opt-out mechanism. Opt-out is an instrument of nudging [8]. It aims to increase the use of EHRs and the acceptance of data sharing by "choice architecture" [9]. Citizens retain their freedom of choice but are nudged toward using the EHR. However, an opt-out mechanism relies on informed citizens to take deliberative action, especially when data sharing can be harmful. The policy of choice architecture and the use of EHR data for health research make clear that health data sharing cannot be reduced to an individual action but must be analyzed in its social context. It raises questions of legitimacy, social inequality, and the diffusion of knowledge among citizens.

This paper focuses on public opinion on EHR data sharing in Germany. It answers three research questions: (1) Is there sufficient willingness in society to share health data that justifies the introduction of an opt-out scheme? (2) Does the willingness to share data differs with socioeconomic status? (3) How can individual willingness to share EHR data be explained?

State of the Art

This section summarizes the research on data sharing and evidence on EHR adoption. International research finds widespread support for the idea of EHRs among citizens and institutions in the health care sector [10-12]. However, "evidence indicates that consumers who have been offered access to PEHRs have not widely used them and the number of active users has remained low" [12]. For example, a recent study shows that in the United States, the number of active users of the EHR is about 26% (2019), despite its increasing integration in medical practice [13]. In Germany, until 2024, according to data from public health insurance, less than 2% of the insured actively requested an EHR [14]. At the same time, only about half of the population is even aware of the EHR [15]. Several studies have shown that interest and use of the EHR show a digital divide. Younger and more digital literate people show a higher probability of using EHRs, while older and low-income groups have less often access [12,16,17]. Regarding socioeconomic factors, lower educational and employment status, as well as belonging to an ethnic minority, have a negative impact on individuals' desire to access and use EHRs [12,18].

Several studies have examined citizens' motivations for using EHRs. The perceived advantages and disadvantages strongly influence the intention to use and the actual use. Key factors here are concerns about data protection and the fear of being discriminated by health insurers or employers on the one hand, and the benefits of better information and experience with the portals or providers on the other [12,19,20]. Data protection concerns are a key factor because health issues are relevant in many areas of daily life [10,12,19]. This can be seen directly when respondents report fears of stigmatization as a result of digitally disseminated health information from the EHR [21,22]. For people with chronic diseases, the balance of advantages and disadvantages seems to be clearly in favor of the advantages. Several studies have reported higher levels of interest in providing data due to the hope of treatment progress [23,24]. For people with chronic diseases, there are already care programs in place that people can benefit from, such as disease management programs [25].

While the research literature stresses that EHRs are potentially an important tool to transform "passive recipients of treatment to empowered consumers" [12], the empirical findings are rather disappointing. Several international studies have shown that the overall use of EHRs by patients remains low [12]. Furthermore, there is little evidence of empowerment effects [12]. Even people who have an EHR often do not know which data are stored in the health record and how to interpret the information [12,26]. When users experience that they do not understand the EHR or the information stored in it, their willingness to use it decreases significantly [23]. Thus, we can expect a significant gap between interest in EHRs and actual use. This gap is likely to be even more pronounced in Germany, with its high standards for data security and complicated routines for accessing the EHR.

Data sharing is an essential part of EHRs. Many EHR systems allow anonymized health data to be used for scientific research. While this can lead to progress in the health system, there is no clear-cut advantage for an individual sharing its data, since anonymization makes individual feedback regarding new health insights hard. That might be the reason, why in public discourse, data sharing is often labeled as "data donation" [27]. Research regarding data sharing has shown that a considerable part of the population (often exceeding 50%) is willing to share its health data [27-30]. There are considerable differences between different countries. For example, the Scandinavian countries show the highest willingness to share health data, while people in Southern Europe are rather skeptical about it [31]. For Germany, recent research suggests that 8 of 10 citizens are willing to share their health data for scientific purposes [15,32,33]. Notably, this high approval rate also seems to hold true for the planned opt-out regulation [33]. However, these high approval rates have to be interpreted with caution, since limited knowledge about health data sharing is widespread and some studies have methodological limitations. Like many other surveys in this area, the high approval rates rely on web-based surveys that have proven to be of less quality regarding the research results compared to face-to-face or telephone interviews [34]. One reason is that they show a higher tendency to produce biased results [35]. Since there is an empirical correlation

between digital literacy and the willingness to share health data, the results are likely to be too optimistic. Furthermore, some questionnaires tend to highlight the positive impacts of data sharing and thereby inflating the approval rate (The question of the survey conducted in Germany was [15,32]: “Assuming that in the future your personal health data, such as your medical history, examination results, X-ray images, etc. can be stored online in a digital health record. Would you agree that your personal health data is used anonymously and free of charge for the purpose of medical research so that in the future diseases can be better recognized and new treatments developed?” And “At present, the use of patient data for medical research requires the consent of patients. However, it is often difficult or even impossible to obtain such consent. Therefore, there are considerations in Germany to legally allow medical research without consent on encrypted patient data after an independent review of the research project. In return, it should be possible for every citizen to simply object to this so-called ‘data donation’ How would you find such a legal regulation ... for publicly funded research?”).

Empirical studies have revealed that the willingness to share data differs among different groups. The findings are comparable to the overall use of EHR: among those with higher age, educational level, and income, a higher willingness to share data can be observed [30,36-39]. Some studies investigate potential individual causes for the willingness to share health data for scientific purposes. These studies find that concerns about confidentiality breaches and data misuse are important factors that prevent data sharing [28,29]. Unsurprisingly, privacy protection, anonymization, and transparency have proven to be key factors for data sharing [29,40]. Furthermore, people are motivated by potential benefits, including more altruistic ones, like advancements in science and public health [29,40]. Another key factor is the use of health data. While international research has shown relatively high public support for the secondary use of health data, this support decreases significantly when shared data can be used by the private sector [41-43].

Another source of insight comes from studies that focus on the social contexts in which data are shared. They have identified 2 other relevant aspects: digital literacy and trust. Health data sharing relies on digital tools. Therefore, people who are more familiar with digital tools may be less reluctant to share data. In addition, technology companies such as Google and Apple have created many services that make use of data sharing in everyday life. Hence, digitally literate people are more involved in data sharing in general. In line with this argument, empirical studies show a higher propensity to share health data among the more digitally savvy people [36,37,40,44,45]. On a more abstract level, it can be argued that trust plays an essential role in health in general [46] and in data sharing in particular. Since the sharing of data is a socially highly indeterminate constellation for social action, trust is a crucial resource. The risks taken by data donors as well as the potential benefits depend strongly on other actors and can hardly be known in advance. For example, the question of anonymizing health data can only be answered in relation to existing knowledge and techniques. The benefits of data sharing are particularly uncertain because there is no guarantee of medical progress, and even if progress does occur, it is unclear whether data donors themselves benefit. Given these theoretical arguments, it is not surprising that empirical studies show an important influence of trust. They can show that trust in the institutions involved in data sharing is relevant, as is general trust in science [15,29,36,37,40,44,45].

Methods

Ethical Considerations

Ethics approval was obtained from the ethics department of the German Aerospace Center (21/23) before the interviews were conducted. Participants gave verbal informed consent to participate in the study. Anonymized interview data were provided to the author by the survey institute. No compensation was given to the interviewees.

Survey Design and Sample

The analysis is based on a dedicated telephone survey (computer-assisted telephone interview [CATI]) about health data conducted between December 7 and 21, 2023. The sampling was carried out by the survey institute drei.fakt (Erfurt) on behalf of the Ernst Abbe University of Applied Sciences Jena. For the study, a representative randomized stratified random sample was drawn based on randomized landline and cell phone numbers (30/70) [47]. Random landline numbers were called based on regional quotas. The mobile phone numbers were sampled and stratified by telephone provider. A total of 11,451 residential numbers were called. In total, 1004 respondents aged 18 years and older participated in an interview. In contrast to many web-based surveys in this area, CATI allows people without internet access to be able to take part in the survey. To ensure the quality of the sample, it was compared with the population for age, gender, and region during the field period. Statistics from the 2019 German microcensus served as a data basis for comparing the distribution of the sample with the distribution of the characteristics in the population. Compared to population statistics, the unweighted sample underrepresents immigrants and slightly overrepresents people with higher education. This selection bias of telephone interviews is well-known from survey research. Therefore, all analyses use a weighting variable in order to adequately represent the population in terms of demographic and socioeconomic characteristics.

Questionnaire

The questionnaire comprised 32 questions regarding various topics about health data sharing. The questionnaire included information on digital skills, the willingness to share personal health data in general, motives for and against data sharing, questions regarding trust and data protection, as well as attitudes toward the EHR and socioeconomic characteristics. The CATI lasted approximately 20 minutes. Most questions were based on a 5-point Likert scale or binary response options and included multiple response categories. The final questionnaire included feedback from 2 separate pretests. In this pretest, the survey instrument was tested for its practical suitability in the field and potential misunderstandings. The full questionnaire and survey data are available in German via the GESIS repository [47].

Data Analysis

Descriptive statistics and multivariate linear regression models are used for analysis. The dependent variable in the multivariate regression models is the attitude toward sharing health data via EHR. People were asked on a 5-point scale (1=yes, definitely and 5=no, definitely not): "Would you be willing to share your health data from your EHR anonymized for medical research purposes?" To minimize potential misunderstandings about the EHR, the question was preceded by a brief explanation of EHRs ("Since January 1, 2021, all people with statutory health insurance can use the EHR via an app of the insurance provider. The EHR contains medical findings and information from previous treatments."). It is important to note that this question is hypothetical, as only a small fraction of people actually had an electronic personal health record at the time of the fieldwork (see Introduction section). However, in order to measure the willingness to share personal health information and to take into account the new legal framework, the question was phrased in a hypothetical manner. This makes it possible to measure the willingness to share extensive and potentially sensitive personal health information rather than actual planned behavior. Therefore, the question is also analyzed for respondents without an EHR in order to reflect overall opinions about the new legislative approach.

Results

The first step is to analyze the approval rates for the use of the German EHR. In general, the population is very open to the EHR. In total, 53% (n=483) of the population who claim to have no EHR yet show interest in using it. At the same time, 8.9% (n=89) of the sample state that they already have an EHR, in comparison to the 1% reported in official statistics. Therefore, it has to be assumed that some people lack a proper understanding of the EHR (see Limitations and Further Research Questions section). Only a minority of 17.1% (n=156) refuse to use it. The proportion of people with no clear opinion (partly partly or do not know) is significantly higher (n=272, 29.9%).

The legislative proposals also introduce the sharing of health data from the EHR for research purposes, as long as individuals do not actively object. Figure 1 summarizes whether respondents would be willing to share their personal health data stored in an EHR. The figure summarizes the attitudes of respondents with and without EHR. Respondents are more cautious about data sharing compared to the interest in using the EHR. In total, 43.4% (n=432) of respondents are willing to share their health data (yes, definitely or rather yes). Nearly 1 in 4 (n=225, 22.6%) respondents oppose data sharing. In addition, a large proportion of respondents are undecided (partly or not at all or do not know). A high interest in using the EHR corresponds empirically with a high willingness to share health data via EHR (no figure). The bivariate correlation is considerably large with 0.56 (Spearman).

The attitude toward health data sharing varies considerably with the data user respondents have in mind (Figure 2). The survey included a question regarding health data sharing with different institutions. It has to be noted that this question did not mention the EHR directly but addressed the willingness to share health data more generally. While 28.7% (n=285) and 18.8% (n=186) do not want to share data with public research institutions or public health insurances, about 38% (n=377) object to share their personal health data with private research institutions.

In the next step, the interest to use the EHR and the willingness to share health data for research purposes are measured for different socioeconomic groups. Table 1 shows the percentage of respondents in each group who intend to use the EHR and how many would agree with data sharing (yes and rather yes). In addition, respondents were asked for their opinions on different consent mechanisms. Figure 1 shows the proportion of people who explicitly reject the opt-out procedure as it will be introduced by law in 2025—compared to those who agree or are undecided (do not know).

Figure 1. Opinions about the use of electronic health record (EHR) and data sharing for research purposes (in percentages). Own calculation and weighted results for the following questions: "Would you like to use the EHR to view and manage your digital health data?" (n=911; nonresponse and people with EHR excluded) and "Would you be willing to share your health data from your EHR anonymized for medical research purposes?" (n=995; nonresponse excluded).

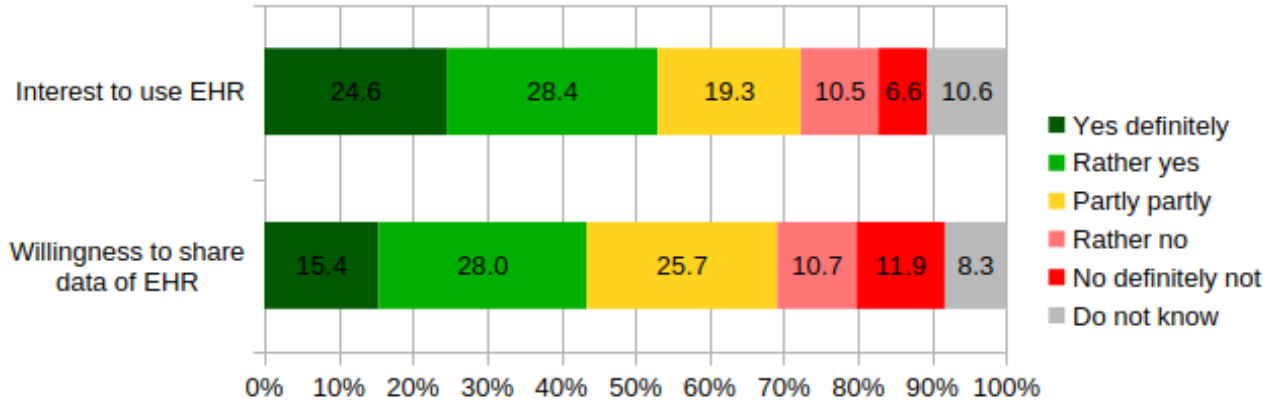


Figure 2. Willingness to share personal health data by data user (in percentages). Own calculation, weighted results for the question, “Would you be willing to share your health data anonymously for medical research purposes with the following institutions?” (nonresponse excluded; n=993).

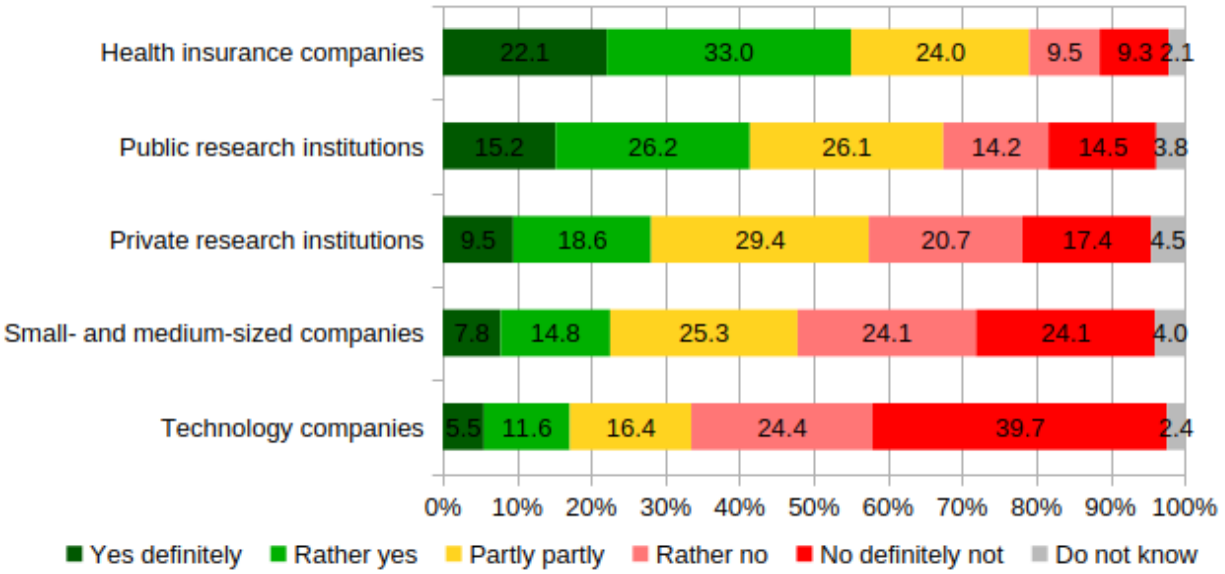


Table . Socioeconomic differences in opinions about electronic health record (EHR)[a].

	Intention to use EHR, n (%)	Willingness to share data of EHR, n (%)	Disagree with opt-out, n (%)
Region			
West Germany	386 (52.1)	350 (43.2)	372 (46.8)
East Germany	97 (57)	82 (44.3)	67 (36)
Gender			
Woman	237 (50.2)	200 (39.1)	228 (45)
Man	245 (56)	232 (47.8)	210 (44.4)
Migrant (not born in Germany)			
No	484 (56.6)	430 (46)	409 (44.5)
Yes	18 (37.3)	18 (33.5)	25 (48.7)
Income quartile			
First (<1250)	67 (48.6)	50 (33.4)	57 (38.5)
Second (<1833)	120 (54.4)	104 (43.9)	90 (38.8)
Third (<2500)	103 (56.5)	97 (49.2)	91 (47.3)
Fourth	134 (61.4)	142 (57.3)	119 (48)
Age (years)			
18 - 39	120 (48.8)	116 (41.2)	121 (44)
40 - 59	164 (52.7)	140 (42.1)	163 (49.5)
Above 60	201 (56.6)	176 (46.5)	154 (41.2)
Educational level			
Low	99 (51.8)	79 (38.6)	70 (35.3)
Medium	182 (52.9)	172 (46.8)	148 (40.6)
High school degree	84 (49.4)	68 (35.9)	98 (51.9)
University degree	118 (57.7)	116 (50.8)	131 (58.1)
Total sample	483 (53)	432 (43.4)	438 (44.7)
Total respondents, n	911	995	980

[a]Own calculations, weighted percentages for columns 1 and 2 (share of a sample with statement yes definitely and rather yes) and column 3 (share of a sample who disagrees opt-out) in response to the question: “The German government is planning to use data from the electronic health records for medical research, eg, for research on new diseases and therapeutic approaches. Which of the following consents do you agree with? ... My data will be shared until I object.” Options to answer: agree or disagree or do not know or nonresponse; nonresponse excluded; do not know is interpreted as no interest to use EHR, not in favor of data sharing and agree with opt-out.

There are clear socioeconomic differences in interest in using the EHR and willingness to share health data, which are similar for both variables. The higher a person’s socioeconomic position (like income or education), the more open respondents are to using the EHR and sharing their personal health data for research purposes. Older people and men also tend to be more open. There are pronounced differences by migration status. The final column summarizes attitudes toward the opt-out process. First of all, it is striking that the procedure implemented by the legislator is explicitly rejected by a significant part of the population. In the entire sample, 44.7% (n=438) say that they are against an opt-out. Additional analyses not presented here show that the remaining respondents either agree (n=400, 40.8%) or have no opinion (n=142, 14.5%). Interestingly, the opt-out model is mainly rejected by people in a higher socioeconomic position. There are hardly any differences with regard to the other socioeconomic variables, except for a lower level of rejection in the eastern part of Germany. Overall, the descriptive statistics suggest that citizens are open to EHR and data sharing but want a high degree of autonomy in dealing with their health data. Additional analyses not shown here indicate that active consent is supported by over 80% (n=788) of respondents.

In the next step, the willingness to share data via the EHR is examined in a multivariate linear regression model (Table 2). The dependent variable is the willingness to share personal health data expressed on a 5-point scale (Figure 1). To prevent an excessive reduction in the sample, people who answered “do not know” were assigned to the “partly partly” category. The findings presented here remain stable even with a smaller sample. The determinants identified in the State of the Art section are added as independent variables to the model. Special

attention is given to the role of trust. The first model includes socioeconomic parameters. As shown in Table 1, the multivariate model confirms socioeconomic differences. However, some more differentiated statements can be made. Controlling for other parameters, there is a curvilinear relationship between age and willingness to share. The positive squared age term indicates that the willingness to share is highest among the very young and the very old. The willingness to share is lowest among people in their 50s. On a 5-point scale, their willingness is about 0.5 points lower than that of 20-year-old people. Income has a positive significant influence on willingness to share, which confirms the descriptive findings. The level of education also has an influence. However, when controlling for income, medium levels of education are associated with the highest willingness to share. Women and people born abroad show no different opinion regarding health data sharing after controlling for income and education. The model also includes a dummy variable indicating the presence of chronic diseases. On the one hand, people with chronic diseases may be at a higher risk of data misuse, but on the other hand, they may be more likely to benefit from digitally supported medical progress. In fact, interest in sharing data appears to be 0.2 points higher for people with chronic physical diseases.

The second model adds comfort with digital technologies. As in comparable studies from other countries, there are significant effects. People who find it easy to use digital technology are more open to sharing their health data. However, the effect disappears when the trust variables are included (model 3). Trust turns out to be a very strong influencing factor. Trust in science has the strongest influence. This is particularly noteworthy because the trust indicator does not specifically address the EHR or the institutions behind it but science as a whole. Controlling for socioeconomic parameters, a person who mistrusts the scientific system has a willingness to share that is more than 2 points lower on a scale of 5 than a person who fully trusts the scientific system. In addition, trust in the big technology firms also has a highly significant influence on the hypothesized direction. The strong influence of trust can be seen directly in the sharp increase in the fit of the model (R^2). The model also includes a proxy for altruistic tendencies in the form of a general willingness to donate (eg, money). The dummy for people who donate on a regular basis is significant. In the last model, motivations and concerns that may be associated with data sharing were included. However, the last model should be interpreted with caution. Because the wording of the questions in the independent variables refers to data sharing, the coefficients may be biased by self-reference (eg, if you care about doing good for society by sharing your data, you will more be willing to share). Nevertheless, the coefficients provide important clues about the motives that drive data sharing. The multivariate results show an interplay between altruistic motives (doing good for society) and cost-benefit considerations (potential benefits for personal and fear of data breaches that expose personal health data to the public).

Table . Willingness to share data of electronic health record for research purposes—linear regression model[a].

	Model 1 (n=828)		Model 2 (n=828)		Model 3 (n=828)		Model 4 (n=828)	
	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value	Coefficient	*P* value
Migrant (0/1[b])	−0.110	.57	−0.116	.54	−0.0190	.92	−0.0431	.83
Gender (reference=woman and 1 nonbinary person)								
Man (0/1)	0.153	.09	0.120	.19	0.0947	.24	0.140	.07
Age (years)	−0.0530[c]	.002	−0.0543[c]	.002	−0.0254	.11	−0.0133	.38
Age² (years)	0.000515[c]	.002	0.000542[c]	.001	0.000293	.06	0.000163	.27
Educational level (reference=medium)								
Low (0/1)	−0.233	.05	−0.217	.07	−0.184	.08	−0.151	.13
High school (0/1)	−0.321[d]	.02	−0.326[d]	.01	−0.319[c]	.01	−0.251[d]	.03
University (0/1)	0.00286	.98	−0.0100	.93	−0.0246	.83	−0.00558	.95
Income[e]	0.000130[c]	.002	0.000128[c]	.002	0.000136[f]	<.001	0.000116[f]	.001
Chronic disease (0/1)	0.201[d]	.03	0.206[d]	.03	0.118	.19	0.0804	.31
Mental illness (0/1)	0.0810	.56	0.0855	.54	0.0441	.73	−0.00517	.96
Rare internet use (0/1)[g]	—[h]	—	−0.229	.09	−0.235	.07	−0.0940	.42
Digital openness[i]	—	—	0.107[d]	.02	0.0330	.42	0.0379	.33
Trust in ...[j]								
Sciences	—	—	—	—	0.417[f]	<.001	0.191[f]	<.001
Big technology companies	—	—	—	—	0.181[f]	<.001	0.0948[d]	.02
Donate regularly (0/1)[k]	—	—	—	—	0.183[d]	.026	0.0474	.54
Motives for data sharing								
Doing good for society[l]	—	—	—	—	—	—	0.206[f]	<.001
Improve my health[m]	—	—	—	—	—	—	0.293[f]	<.001
Afraid of data breach (0/1)[n]	—	—	—	—	—	—	−0.206[c]	.007
Constant	4.307[f]	<.001	3.927[f]	<.001	1.448[c]	.002	0.571	.19

[a]Own calculations, weighted results. Model 1: R^2=0.059; model 2: R^2=0.072; model 3: R^2=0.270; and model 4: R^2=0.430.

[b]0/1 points to dummy variables.

[c]*P*<.01.

[d]*P*<.05.

[e]Net household equivalence income, imputed (5 times).

[f]*P*<.001.

[g]Once a week or less.

[h]Not applicable; regression model with fewer covariates.

[i]Using digital hardware and software feels easy for me (1.5=agree totally).

[j]Trust (1.5=trust totally) in technology companies like Google, Facebook, Apple or in sciences.

[k]Donate for charity on a regular basis (monthly or yearly).

[l]Subjective importance that with data sharing I am doing good for society (1.5=very important).

[m]Subjective importance that with data sharing I profit regarding my personal health.

[n]I am worried that my data will be made public (0=disagree or 1=agree 1).

Discussion

Principal Findings

The results of this study shed light on attitudes toward the use and sharing of EHRs for research purposes in Germany, particularly in the context of the new legislation based on the opt-out mechanism. Overall, the study reveals a nuanced picture of the German population's attitudes. The empirical results show an openness to EHRs (about half of the nonusers express a desire to use their EHR, signaling significant potential for increased adoption). At the same time, a significant share of the population remains skeptical or undecided. The overall openness is in line with international research regarding interest in EHR [12]. Given these high levels of interest, the introduction of an opt-out model does not seem unjustified. However, especially a large fraction of undecided people need to be properly informed about the consequences of having their health data stored electronically. While there is general support for sharing health data, with 43.4% (n=432) of respondents indicating a willingness to do so, a significant portion (n=225, 22.6%) remain opposed, and a large number are undecided. The results indicate a significantly lower willingness to share health data than has been measured in comparable studies in Germany [15,33]. Possible explanations for these differences are (1) different timing: other studies collected their data at a time when there was no legal framework for health data sharing and therefore risks, etc, had not yet been widely discussed in public. (2) Methodological: this study is based upon a telephone survey rather than web-based and explicitly allows people to have no specific attitude by including categories like refuse to answer, do not know, and neither nor. Last but not least, the comparatively low willingness to share health data in this study compared to international research [27,28] may be due to the emphasis on the personal character of the data potentially shared. The ambivalent results highlight the need for careful consideration of how the opt-out mechanism is implemented and communicated. Although there is openness in principle, "soft paternalism" must be viewed critically precisely because of the high proportion of undecided citizens. In contrast to the principle of informed consent, in which the processes and risks of data sharing have to be explained before citizens allow data sharing, an opt-out solution presupposes that the population is sufficiently informed in other ways. This becomes even more clear after considering attitudes toward the opt-out mechanism, which 44.7% (n=438) disapprove. Especially people who are positive about EHRs and the sharing of their health data reject the opt-out mechanism proposed by the legislator more often. Presumably, "choice architecture" is seen as too much interference in their own decisions.

The study also confirms a digital divide, as socioeconomic factors significantly influence both the willingness to use EHRs and the willingness to share personal health data. The results are in line with international research findings [16,30]. Higher income, education, and digital literacy are associated with greater openness, while those with lower socioeconomic status are more reluctant. However, socioeconomic factors are not the most important in explaining openness to data sharing. Trust plays a critical role in determining the willingness to share health information. While other studies have pointed to the role of trust as a significant factor [15,40,44,45], this study shows that institutional trust has a decisive influence compared to other factors. The influence of socioeconomic variables, as well as other factors such as digital literacy, diminishes when trust variables are taken into account, highlighting trust as a fundamental element in the acceptance of digital health initiatives. The observed socioeconomic disparities reinforce the need for targeted interventions and information to ensure equitable access and participation in digital health.

Policy Implications

The paradigm shift to an opt-out system is expected to increase participation rates because it relies on passive inclusion rather than active consent. However, this comes at the cost of potentially creating a population of uninformed and passive citizens who are enrolled in the system without fully understanding the implications. This raises critical concerns about consumer autonomy and challenges the notion of the autonomous consumer who actively engages with—and consents to—data-sharing practices. With a large portion of the population still undecided about EHR adoption and data sharing, the need for population-wide education becomes even more apparent. Ideally, all citizens, regardless of their socioeconomic background, should be able to make informed and autonomous decisions about their participation. The study also shows the complex mechanisms of the digital divide, with those in higher socioeconomic positions being more open to the adoption of EHRs but also more likely to oppose the opt-out system. This suggests that the paradigm shift is perceived as a restriction of autonomy. Those with greater resources and digital literacy want to make informed choices and therefore resist mechanisms that limit their freedom of choice. While those with greater resources are likely to be able to opt out if they perceive the risks of data sharing to be too high, they could trigger serious political resistance to the opt-out mechanism. In contrast, individuals with lower socioeconomic status may be less able to navigate these systems and therefore are more vulnerable to passive inclusion.

The transition to an opt-out system for EHRs in Germany marks a significant policy shift with the potential to transform the health sector. However, for this shift to be successful, it is essential to address the underlying concerns of the population, particularly around data security, autonomy, and the digital divide. Continuous monitoring of the implementation of the opt-out system will be crucial to keep and build trust within the population.

Limitations and Further Research Questions

In contrast to web-based surveys, which are often used in EHR research, the CATI method used here avoids many problems of self-reporting, sample bias, and inference [48,49]. However, the quality of data collected through telephone interviews also experiences low response rates. It seems highly plausible that nonresponse is not random. People with lower levels of trust and willingness are supposedly less likely to participate in the survey. Thus, despite the weighting procedure applied, the sample may be biased toward more trusting people. This problem is common to other data used in comparable research.

Another limitation concerns the correct understanding of the EHR in Germany. During the field phase, little more than 1% of the insured population opted for the EHR. It is possible that many respondents do not have a proper understanding of the EHR or have no knowledge at all. The survey provides some evidence in this direction. A question about knowledge of the EHR reveals that 13.9% (n=139) have never heard of it. For this group, the evaluation of data sharing is probably less substantiated. However, at the same time, a large fraction says that they have heard of the EHR (n=720, 71.7%) or even use it (n=89, 8.9%). On the one hand, these results support the assertion that a large proportion of the sample has a basic understanding of the EHR. During the field phase, there was also an extensive media discourse about the EHR, as the Gesundheitsdatennutzungsgesetz was being discussed in parliament in December 2023. On the other hand, the self-reported use of the EHR does not match the official statistics. Some respondents probably confused the EHR with other apps provided by their health insurance. This problem of overreporting has also been observed in other studies [15].

The study focuses on the interest in using the EHR and the willingness to share health data. At the time of the field phase, EHR adoption was very low, and a legal framework for health information exchange was in its formative stages. Research has already shown that there is a gap between attitudes and actual behavior [12,42]. So the results have to be interpreted with some caution. In addition, the high proportion of undecided respondents indicates that reported attitudes are not very salient and therefore susceptible to survey effects. However, other studies share this disadvantage. Moreover, the consequences of limited interest in EHR do not only affect surveys but also have far-reaching practical implications, especially when using opt-out regulations.

Future research should focus on the gap between interest in using the EHR and actual behavior. With a focus on Germany, the potential burdens of adoption should be analyzed in more detail. With regard to trust, the potential ways to build trust within the health care system without relying too much on general institutional trust also point to new research directions. Furthermore, there are interesting research questions regarding self-responsibility in a system that becomes largely based on passive consent (opt-out).

Acknowledgments

This work was supported by the German Federal Ministry of Education and Research and the European Union—Next Generation (grant 16KISA015). The funded project is called AVATAR ("Anonymisierung persönlicher Gesundheitsdaten durch Erzeugung digitaler Avatare") and consists of several research institutions that aim to develop new methods for anonymizing patient data for research. The funding agency had no role in the design, conduct, analysis, interpretation, or approval of the study. The content is the sole responsibility of the author. The author would like to thank Elias Kühnel, Julia Berghäuser, Sven Bock, and Anna Weber for their thorough work on the questionnaire.

Conflicts of Interest

None declared.

References

1. Elektronische Patientenakte bringt Entbürokratisierung und entlastet Ärzte und Pflegekräfte [Web page in German]. Bundesministerium für Gesundheit. URL: https://www.bundesgesundheitsministerium.de/presse/interviews/interview/spiegel-10-11-23-epa-und-ki [accessed 2025-03-05]
2. Digitalisierung für Gesundheit: Ziele und Rahmenbedingungen eines dynamisch lernenden Gesundheitssystems [Web page in German]. Gutachten. 2021. URL: https://www.svr-gesundheit.de/fileadmin/Gutachten/Gutachten_2021/SVR_Gutachten_2021.pdf [accessed 2025-04-08]
3. Strotbaum V, Pobiruchin M, Schreiweis B, Wiesner M, Strahwald B. Your data is gold—data donation for better healthcare? it - Information Technology 2019 Oct 25;61(5-6):219-229. [doi: 10.1515/itit-2019-0024]
4. Lupton D. Digital Health: Routledge; 2018.
5. Caspar J. Wir Datensklaven [Book in German]: Econ; 2023.
6. Matzat U, van Ingen E, Maasen S, Passoth JH. Social inequality and the digital transformation of western society: what can stratification research and digital divide studies learn from each other? In: Maasen S, Passoth JH, editors. Soziologie des Digitalen—Digitale Soziologie: Nomos; 2020:381-397.
7. Merkel S, Hess M. The use of internet-based health and care services by elderly people in Europe and the importance of the country context: multilevel study. JMIR Aging 2020 Jun 3;3(1):e15491. [doi: 10.2196/15491] [Medline: 32490837]
8. Thaler RH, Sunstein C. Nudge Improving Decisions about Health, Wealth and Happiness: Penguin; 2009.

9. Sunstein CR. Choosing Not to Choose: Understanding the Value of Choice: Oxford University Press; 2015.
10. Hoerbst A, Kohl CD, Knaup P, Ammenwerth E. Attitudes and behaviors related to the introduction of electronic health records among Austrian and German citizens. Int J Med Inform 2010 Feb;79(2):81-89. [doi: 10.1016/j.ijmedinf.2009.11.002] [Medline: 20031482]
11. Kruse CS, Stein A, Thomas H, Kaur H. The use of electronic health records to support population health: a systematic review of the literature. J Med Syst 2018 Sep 29;42(11):214. [doi: 10.1007/s10916-018-1075-6] [Medline: 30269237]
12. Crameri KA, Maher L, Van Dam P, Prior S. Personal electronic healthcare records: what influences consumers to engage with their clinical data online? A literature review. Health Inf Manag 2022 Jan;51(1):3-12. [doi: 10.1177/1833358319895369] [Medline: 31918578]
13. Zheng H, Jiang S. Frequent and diverse use of electronic health records in the United States: a trend analysis of national surveys. Digit Health 2022;8:20552076221112840. [doi: 10.1177/20552076221112840] [Medline: 35832476]
14. Seifert A. Nur wenige Deutsche nutzen bislang die Elektronische Patientenakte [Web page in German]. mdr. 2024. URL: https://www.mdr.de/nachrichten/deutschland/gesellschaft/wenige-elektronische-patientenakten-bisher-100.html [accessed 2025-03-05]
15. Haug S, Schnell R, Raptis G, Dotter C, Weber K. Wissen und Einstellung zur Speicherung und Nutzung von Gesundheitsdaten: Ergebnisse einer Bevölkerungsbefragung [Article in German]. Z Evid Fortbild Qual Gesundhwes 2024 Mar;184:50-58. [doi: 10.1016/j.zefq.2023.11.001]
16. Kim EH, Kim Y. Digital divide: use of electronic personal health record by different population groups. Presented at: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology; Aug 31 to Sep 4, 2010; Buenos Aires, Argentina. [doi: 10.1109/iembs.2010.5626732]
17. Halmdienst N, Pruckner GJ, Winter-Ebmer R. Complexities of health and acceptance of electronic health records for the Austrian elderly population. Eur J Health Econ 2023 Feb;24(1):53-66. [doi: 10.1007/s10198-022-01451-z] [Medline: 35287235]
18. Paccoud I, Baumann M, Le Bihan E, et al. Socioeconomic and behavioural factors associated with access to and use of personal health records. BMC Med Inform Decis Mak 2021;21:18. [doi: 10.1186/s12911-020-01383-9]
19. Henkenjohann R. Role of individual motivations and privacy concerns in the adoption of German electronic patient record apps—a mixed-methods study. Int J Environ Res Public Health 2021 Sep 10;18(18):9553. [doi: 10.3390/ijerph18189553] [Medline: 34574475]
20. Griesser A, Bidmon S. A process related view on the usage of electronic health records from the patients' perspective: a systematic review. J Med Syst 2022 Dec 29;47(1):2. [doi: 10.1007/s10916-022-01886-0] [Medline: 36580132]
21. Stablein T, Hall JL, Pervis C, Anthony DL. Negotiating stigma in health care: disclosure and the role of electronic health records. Health Sociol Rev 2015 Sep 2;24(3):227-241. [doi: 10.1080/14461242.2015.1078218]
22. Kalckreuth N, Prümper AM, Feufel MA. The influence of health data on the use of the electronic health record (EHR): a mixed methods approach. 2023 Presented at: AMCIS 2023 Proceedings; Aug 10, 2023; Panama.
23. Eriksson-Backa K, Hirvonen N, Enwald H, Huvila I. Enablers for and barriers to using My Kanta— a focus group study of older adults' perceptions of the National Electronic Health Record in Finland. Inform Health Soc Care 2021 Dec 2;46(4):399-411. [doi: 10.1080/17538157.2021.1902331] [Medline: 33787438]
24. Greenberg AJ, Falisi AL, Finney Rutten LJ, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017;19(6):e188. [doi: 10.2196/jmir.7417]
25. Nolte D. Vom Daten-Grab zum Daten-Schatz? Strukturierte Behandlungsprogramme (Disease Management Programme) für chronisch Kranke. Zugluft 2021(2):82-87 [FREE Full text]
26. Caine K, Kohn S, Lawrence C, Hanania R, Meslin EM, Tierney WM. Designing a patient-centered user interface for access decisions about EHR data: implications from patient interviews. J Gen Intern Med 2015 Jan;30(Suppl 1):S7-16. [doi: 10.1007/s11606-014-3049-9] [Medline: 25480719]
27. Skatova A, Goulding J. Psychology of personal data donation. PLOS ONE 2019;14(11):e0224240. [doi: 10.1371/journal.pone.0224240] [Medline: 31747408]
28. Howe N, Giles E, Newbury-Birch D, McColl E. Systematic review of participants' attitudes towards data sharing: a thematic synthesis. J Health Serv Res Policy 2018 Apr;23(2):123-133. [doi: 10.1177/1355819617751555]
29. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan;48(1):3-13. [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
30. Hirst Y, Stoffel ST, Brewer HR, Timotijevic L, Raats MM, Flanagan JM. Understanding public attitudes and willingness to share commercial data for health research: survey study in the United Kingdom. JMIR Public Health Surveill 2023;9:e40814. [doi: 10.2196/40814]
31. Kühnel E, Wilke F. Potentiale der Digitalisierung heben: Die Bereitschaft zur Spende von Gesundheitsdaten in Deutschland und Finnland [Chapter in German]. In: Heckes KT, Lorke M, Siegler M, Wrona KJ, editors. Soziotechnische Transformationen im Sozial- und Gesundheitswesen: kollaborativ, divers, barrierefrei und sozialräumlich: Juventa; 2025.

32. Richter G, Borzikowsky C, Hoyer BF, Laudes M, Krawczak M. Secondary research use of personal medical data: patient attitudes towards data donation. BMC Med Ethics 2021 Dec 15;22(1):164. [doi: 10.1186/s12910-021-00728-x] [Medline: 34911502]
33. Richter G, Trigui N, Caliebe A, Krawczak M. Attitude towards consent-free research use of personal medical data in the general German population. Heliyon 2024 Mar 30;10(6):e27933. [doi: 10.1016/j.heliyon.2024.e27933] [Medline: 38509969]
34. Lee H, Kim S, Couper MP, Woo Y. Experimental comparison of PC web, smartphone web, and telephone surveys in the new technology era. Soc Sci Comput Rev 2019 Apr;37(2):234-247. [doi: 10.1177/0894439318756867]
35. Milton AC, Ellis LA, Davenport TA, Burns JM, Hickie IB. Comparison of self-reported telephone interviewing and web-based survey responses: findings from the Second Australian Young and Well National Survey. JMIR Ment Health 2017 Sep 26;4(3):e37. [doi: 10.2196/mental.8222] [Medline: 28951382]
36. Moon LA. Factors influencing health data sharing preferences of consumers: a critical review. Health Policy Technol 2017 Jun;6(2):169-187. [doi: 10.1016/j.hlpt.2017.01.001]
37. Karampela M, Ouhbi S, Isomursu M. Connected health user willingness to share personal health data: questionnaire study. J Med Internet Res 2019 Nov 27;21(11):e14537. [doi: 10.2196/14537] [Medline: 31774410]
38. Weng C, Friedman C, Rommel CA, Hurdle JF. A two-site survey of medical center personnel's willingness to share clinical data for research: implications for reproducible health NLP research. BMC Med Inform Decis Mak 2019 Apr 4;19(Suppl 3):70. [doi: 10.1186/s12911-019-0778-z] [Medline: 30943963]
39. Broekstra R, Aris-Meijer JL, Maeckelberghe ELM, Stolk RP, Otten S. Motives for withdrawal of participation in biobanking and participants' willingness to allow linkages of their data. Eur J Hum Genet 2022 Mar;30(3):367-377. [doi: 10.1038/s41431-021-00997-5] [Medline: 34803164]
40. Mählmann L, Schee gen. Halfmann S, von Wyl A, Brand A. Attitudes towards personal genomics and sharing of genetic data among older Swiss adults: a qualitative study. Public Health Genomics 2018 Apr 10;20(5):293-306. [doi: 10.1159/000486588]
41. Braunack-Mayer A, Fabrianesi B, Street J, et al. Sharing government health data with the private sector: community attitudes survey. J Med Internet Res 2021 Oct 1;23(10):e24200. [doi: 10.2196/24200] [Medline: 34596573]
42. Benevento M, Mandarelli G, Carravetta F, et al. Measuring the willingness to share personal health information: a systematic review. Front Public Health 2023;11:1213615. [doi: 10.3389/fpubh.2023.1213615] [Medline: 37546309]
43. Kühnel E, Wilke F. Health data sharing in Germany: individual preconditions, trust and motives. Front Public Health 2025;13:1538106. [doi: 10.3389/fpubh.2025.1538106] [Medline: 39975784]
44. Weitzman ER, Kaci L, Mandl KD. Sharing medical data for health research: the early personal health record experience. J Med Internet Res 2010 May 25;12(2):e14. [doi: 10.2196/jmir.1356] [Medline: 20501431]
45. Maiorana A, Steward WT, Koester KA, et al. Trust, confidentiality, and the acceptability of sharing HIV-related patient data: lessons learned from a mixed methods study about Health Information Exchanges. Implement Sci 2012 Apr 19;7:34. [doi: 10.1186/1748-5908-7-34] [Medline: 22515736]
46. Meyer S, Ward P, Coveney J, Rogers W. Trust in the health system: an analysis and extension of the social theories of Giddens and Luhmann. Health Sociol Rev 2008 Aug;17(2):177-186. [doi: 10.5172/hesr.451.17.2.177]
47. Wilke F, Kühnel E. Eine Befragung zur Bereitschaft zum Teilen von Gesundheitsdaten in Deutschland—Einflussfaktoren, Präferenzen und Motive. : GESIS, Köln; 2024.
48. McPhee C, Barlas F, Brigham N, Darling J, Dutwin D, et al. Data quality metrics for online samples: considerations for study design and analysis. : American Association for Public Opinion Research; 2022 URL: https://aapor.org/wp-content/uploads/2023/02/Task-Force-Report-FINAL.pdf [accessed 2025-04-01]
49. Pokorny S, Hirndorf D. Online, offline oder beides? Umfragemethoden im Praxistest [Article in German]. Analysen Argumente 2024(519):1-13 [FREE Full text]

Abbreviations

CATI: computer-assisted telephone interview
EHR: electronic health record

Edited by A Kushniruk; submitted 23.08.24; peer-reviewed by S Haug, SH Sung; revised version received 04.02.25; accepted 04.03.25; published 15.04.25.

Please cite as:
Wilke F
The Impact of Trust and the Role of the Opt-Out Mechanism in Willingness to Share Health Data via Electronic Health Records in Germany: Telephone Survey Study
JMIR Hum Factors 2025;12:e65718
URL: https://humanfactors.jmir.org/2025/1/e65718
doi:10.2196/65718

Screening Workers for Occupational Exposure to Respirable Crystalline Silica: Development and Usability of an Electronic Data Capture Tool

Fiona Hore-Lacy[1,2]; Christina Dimitriadis[1]; Ryan F Hoy[1,2]; Javier Jimenez-Martin[1]; Malcolm R Sim[1]; Jane Fisher[1]; Deborah C Glass[1]; Karen Walker-Bone[1]

[1]School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia
[2]Department of Respiratory Medicine, Alfred Health, Melbourne, Australia

Corresponding Author:
Fiona Hore-Lacy
School of Public Health and Preventive Medicine, Faculty of Medicine, Nursing and Health Sciences, Monash University, Melbourne, Australia

Abstract

Background: Cases of the occupational lung disease silicosis have been identified in workers processing artificial stone in the stone benchtop industry (SBI). In the Australian state of Victoria, the Regulator commissioned a screening program for all workers in this industry.

Objective: To facilitate systematic data collection, including high-quality exposure assessment, an electronic data capture tool (EDCT) was developed.

Methods: A multidisciplinary team developed an EDCT using Research Electronic Data Capture (REDCap; Vanderbilt University). The needs of the EDCT were (1) data entry by multiple clinicians and the workers attending for screening and (2) systematic collection of data for clinical and research purposes. The comprehensibility and utility of the tool were investigated with a sample of workers, and the EDCT was subsequently refined.

Results: The EDCT was used in clinical practice, with capacity for data extraction for research. Testing of comprehension and utility was undertaken with 15 workers, and the refined version of the Occupational Silica Exposure Assessment Tool (OSEAT) was subsequently developed.

Conclusions: The refined OSEAT has been determined to be comprehensible to workers and capable of collecting exposure data suitable for assessment of risk of silicosis. It was developed for workers in the SBI in Australia and is adaptable, including translation into other languages. It can also be modified for SBI workers in other countries and for use by workers from other industries (mining, construction) at risk of silica exposure, including in lower-income settings.

(JMIR Hum Factors 2025;12:e64111) doi:10.2196/64111

KEYWORDS

silicosis; occupational history; electronic data capture tool (EDCT); REDCap; occupational respiratory screening; occupational hazard; exposure; silica; fibrotic lung disease; lung disease; respirable crystalline silica; mining; construction; workers; occupational lung disease; occupational; Australia; screening

Introduction

Silicosis is an incurable, potentially life-threatening, form of fibrotic lung disease caused by inhalation of respirable crystalline silica (RCS) [1]. The disease has been recognized globally for over 100 years, and lung disease screening is recommended for high-risk industries, including mining and construction [2]. Cases of silicosis were identified in 2010 among workers in the stone benchtop (countertop) industry (SBI) working with artificial stone (AS; Textbox 1) [3-12]. Subsequently, a number of cases of artificial stone-associated silicosis were diagnosed in Australia [3,13]. AS has a very high crystalline silica content, often over 90% [14]. Processing AS by drilling, polishing, cutting, or grinding generates fine particles of dust containing RCS, which can cause silicosis when inhaled [15].

Textbox 1. Timeline of artificial stone (AS) and silicosis in Australia

Early 2000s: AS introduced to Australia
2010: First case of silicosis associated with AS reported in Italy
2015: First case of silicosis associated with AS reported in Australia (conference abstract)
2017: First case of silicosis associated with AS reported in Australia
2019: Screening program of stone benchtop industry workers begins in Victoria, Australia (paper-based data collection)
2021: Screening program first incorporates electronic data capture tool (EDCT)
2021-2023: Refinement of EDCT informed by data cleaning and evaluation study with workers

An investigation of the effects of RCS exposure in the SBI was commissioned by the Victorian regulator, WorkSafe Victoria [16], and developed into a screening program by Monash University. It included (1) exposure assessment from a detailed occupational history; (2) collection of respiratory symptoms; (3) recording of investigations including spirometry, chest x-ray, and high-resolution CT of the chest; (4) screening for comorbidities associated with silica exposure, including autoimmune diseases and tuberculosis [4,17]; and (5) a mental health instrument.

The initial paper-based questionnaire was developed by a multidisciplinary team, including respiratory physicians, an occupational hygienist, and occupational physicians [16]. Simplicity was prioritized, as many industry workers were born outside Australia and spoke English as an additional language [18].

Up to 6 jobs in the SBI could be recorded in the occupational history. The proportion of time spent on specified *tasks* in each job and the proportion of time spent on dry cutting of stone and working near someone doing dry cutting were recorded. *Exposure control measures* (ventilation and respirator use) were identified for each job. Other information collected included the country, start and (if relevant) finish date; days per week worked; number of people in the organization; and type of stone predominantly worked with (AS or natural stone). Other silica-associated occupations (eg, mining, quarrying) and any non-occupational activities that involved dust exposure (including hobbies and home repairs, eg, tiling, plastering) were also recorded.

Data were collected from multiple users, including respiratory physicians, multidisciplinary team, workers, and administrative staff, and capture all the elements listed earlier. The data were cleaned and entered into an electronic data capture tool (EDCT) held on the secure REDCap platform [19,20]. The data were used for both clinical and research purposes. In 2021, screening was centralized and carried out at a single site, which led to a need for direct data entry to the EDCT by the worker.

Exposure calculations from the occupational data have been used to identify roles within the SBI with greater RCS exposure, such as factory machinists and installers, and that exposure intensity and cumulative exposure were associated with dyspnea and radiological abnormalities consistent with silicosis [21]. The screening program data has also been used to describe the numbers of cases of silicosis diagnosed to date [18], the rates and determinants of psychological distress, and the psychometric properties of the mental health instrument [22,23].

The aim of this study was to describe the development of the EDCT, present its refined content, and describe the results of an audit of its clinical utility undertaken with a sample of workers.

Methods

Overview

The original team reviewed the exposure questionnaire, which overall had been well understood, and identified items that required substantial data cleaning. The team redeveloped the questions, which included, for example, adding illustrations of dry and wet cutting examples and the types of respirator and ventilation options that were sourced from workplace health and safety organizations [24-26]. Further, additional optional responses were added from free text replies, for example, water jet cutting.

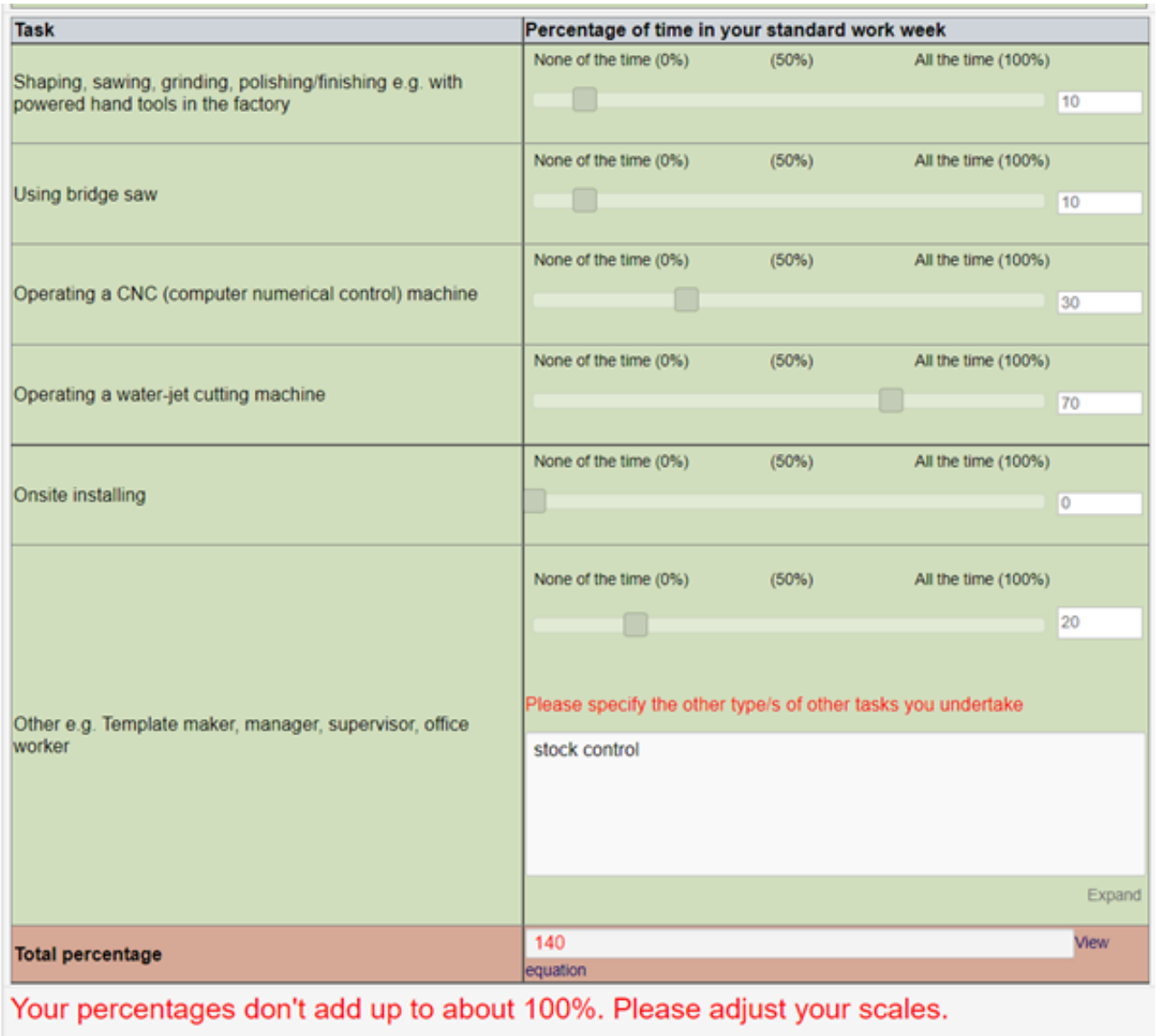
In the first draft questionnaire, participants were asked to apportion their tasks (Figure 1).

The percentages seldom added to 100%, as shown in the example in Figure 1, so in the revised questionnaire, a sliding bar was provided that provided visual input of the proportions (Figure 2). A pop-up trigger was included if the task proportions were out of range, as shown in Figure 2.

Figure 1. Original paper questionnaire asking workers to estimate proportions of work time spent doing specific tasks in their workplace. CNC: Computer Numerical Control____.

15.	What do you do in this job? (Tick all that apply)	
TASK		Estimate percentage of time in a typical week undertaking this task (e.g. 10%)
Shaping e.g. with powered hand tools	☑	20
Sawing e.g. with bridge saw	☑	50
Using CNC machine	☐	
Polishing/Finishing	☑	50
General labouring	☑	50
Maintenance	☑	40
Cleaning the tools, surfaces and/or work space	☑	50
Onsite Installing	☑	100
Other eg Template maker, manager, supervisor, office worker	☐	
If other, specify:_____		

Figure 2. EDCT (electronic data capture tool) version of the task estimation section of the occupational history with a total that adds to a proportion of >100% of the time and the warning message provided to the worker.



Prior to attending for screening, workers were emailed a link to the EDCT containing the revised questions in order to

complete their occupational history. The final page completed by workers included optional electronic consent for sharing data with Monash University.

Other data collected during screening included medical history, smoking status, respiratory symptoms, physical examination findings, diagnosis, return-to-work assessment, and results of all relevant investigations, including chest x-ray, high-resolution CT, pathology, and spirometry.

In 2023, an investigation of the comprehensibility, feasibility, and face validity of the occupational history section of the EDCT was completed with workers.

A total of 15 workers participated in the investigation, which was conducted between February and May 2023. Workers were interviewed about their responses on the EDCT, using a pre-developed proforma to prompt feedback. They were asked: "Were there any words you did not know?" "Did you understand what was meant by 'dry work'" "Are there any other tasks in your workplace that expose you to dust?" and "Did the list of ventilation options include what you use in your workplace?" For all questions, respondents could reply yes or no and provide additional comments. If a worker reported no dry cutting in their current or most recent job but had exposure to dry cutting from previous jobs, they were asked to respond to questions referring to their former job.

Ethical Considerations

Workers were eligible to participate if they had completed their occupational history using the EDCT, did not require an interpreter for their appointment, and had provided electronic consent to share data with Monash University. All participants were provided with a written information sheet and asked to provide verbal consent for participation. Responses were deidentified, and workers did not receive compensation for their participation. Approval was granted by the Alfred Hospital Research Ethics Committee as a substudy of project 292/21.

Results

The demographics of the workers are presented in Table 1. Most participants were male, and they covered a range of ages and years of employment in the SBI. They included machinists, installers, and office workers.

Table . Demographics of participants.

Characteristics		Values
Males, n (%)		14 (93)
Age (years), mean (SD)		38.1 (10.9)
Years in stone benchtop industry, mean (SD)		9.6 (7.85)
Born in Australia, n (%)		5 (33)
Language other than English spoken at home, n (%)		6 (40)
Most recent SBI job held, n		
	Director	3
	Installer	4
	Stonemason	5
	Foreman	2
	Other	1

All participants were asked to report about their *comprehension* of an introductory statement, and one worker commented that it took a while to read and understand, and did not feel like he understood it fully. All workers reported that they comprehended what was meant by "dry work." When asked about *tasks* that exposed them to dust that were not already listed on the EDCT, 2 workers identified new relevant tasks: emptying bins containing benchtop fragments and cleaning of the final benchtop product, onto which the dust-containing water used in wet cutting had dried. In total, 4 workers reported that their *ventilation* option was not included in the list on the EDCT, nor was it pictured. However, after discussion, the alternative options they were describing were "air conditioning," "garage door," "no ventilation," and "ventilation in the wall," all of which were listed.

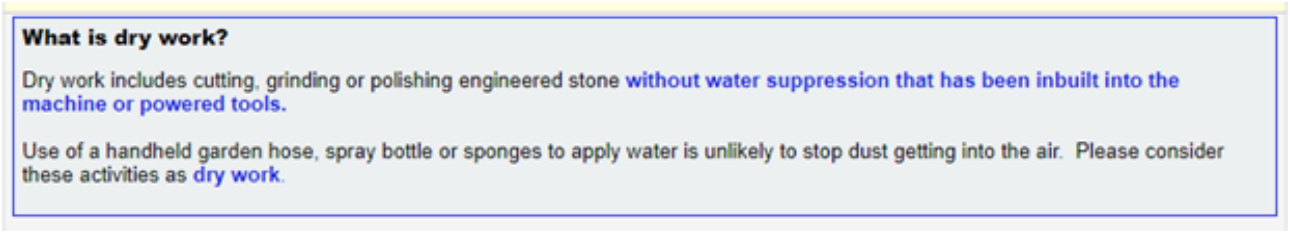
For *respirators*, 14 workers identified the type they used from the descriptions and pictures in the EDCT, and the only worker that did not see their device described use of one similar to that depicted. Of 8 workers who were asked if they could easily remember and estimate the percentage of time they spent wearing a respirator, 7 responded in the affirmative. One worker had worn his respirator for 6 hours out of an 8-hour shift (approximately 75%), but had estimated that he wore it for 35% of the day. One worker commented that he wore his respirator all day, regardless of the task, whilst another pointed out that each job was different, with some jobs being "perfect" (ie, not requiring any adjustments), whereas others required adjustments onsite, adding to the difficulty of responding to this question accurately.

If a worker reported dry cutting in their current job, they were asked whether or not practices around dry cutting had changed (this was because dry cutting without suitable protection had been officially banned recently in Australia). Of the 11 workers who completed this question, 10 understood the question and were able to complete it accurately, but one worker expressed some confusion around the wording of the question. Workers

also described measures other than ventilation and respirators that their employer had introduced including changing clothes at work and using water systems for dust suppression when loading stone.

Subsequently, modifications were made to improve accessibility for the workers: simplification of the language used in descriptive statements and instructions, modification of the color scheme to improve readability, and addition of commonly reported responses (eg, "home maintenance," "tiling," and "plastering") as a prompt on the non-occupational (eg, hobby) dust exposure history section. One of the simplified statements was an introductory statement that defined "dry work" (Figure 3). An image accompanied the text with examples of dry and wet cutting [26]:

Figure 3. Refined EDCT for clarification of the definition of dry work.



The current version of the Occupational Silica Exposure Assessment Tool from the EDCT is presented in Multimedia Appendix 1.

Discussion

Principal Findings

Since 2019, over 1000 SBI workers have undergone screening for silicosis in Victoria, Australia, through a protocolized screening program. In this study, we described the development of the screening questionnaire into an EDCT, the Occupational Silica Exposure Assessment Tool, and have described how it was refined as a result of our experiences and after assessment of its acceptability and comprehension among this worker population. The results of the investigation suggested that the usability and comprehension of the refined EDCT are acceptable among English-speaking workers.

The benefits of EDCTs for improving patient care [27]; improving accuracy of data collection compared to paper methods [28]; and facilitating data collection from multiple users, including patients and health care providers [29], have been established in many settings. Moreover, EDCTs are able to capture and retain large volumes of data, maximizing cost- and time-efficiency in clinical and research settings [29,30]. As demonstrated, the development of this EDCT has already provided all of these benefits and enabled us to create a streamlined and efficient screening program for workers in the artificial stone benchtop industry.

Benchtops made from AS are a popular kitchen product globally, and there are concerns that cases of silicosis among workers who produce them are underreported in the literature [31,32]. Globally, silica deaths were estimated to be more than 12.9 thousand in 2019 [33], and the highest rates were recorded in low- and middle-income countries [33,34]. Silicosis has been seen in a range of industries, including construction, jewelry production, quarrying, tunneling, dental material manufacturing, denim jean production, and ceramic and pottery manufacturing [35]. There is therefore an urgent need for occupational screening for large numbers of workers exposed to RCS, for which reliable instruments are needed. The Occupational Silica Exposure Assessment Tool can be deployed in settings in which workers are exposed to RCS, whether for workers in the SBI or modified for other occupational settings.

In addition to assessing SBI workers at risk of silicosis, the data collected from the Occupational Silica Exposure Assessment Tool (OSEAT) can be used to estimate an individual's level of RCS exposure. In previous work, occupational history data collected using the OSEAT were used to group SBI workers by extent of silica exposure, using a combination of the proportion of time working with AS and the proportion of time spent dry cutting [21]. Both cumulative exposure and exposure intensity were found to be associated with symptoms of dyspnea and chest x-ray abnormalities [21]. This illustrates the ready utility of having added the e-consent function to the EDCT within REDCap, facilitating extracting data from the OSEAT to use for research purposes.

One consideration when introducing a REDCap-based EDCT in a clinical setting is its reliance on workers having adequate internet connection. This has been a limiting factor for the utilization of similar instruments, in which open-source software that was not reliant on internet connection was preferred [28,36]. In some resource-poor settings, REDCap was the preferred platform [37,38], and REDCap has now developed a mobile app that can be used offline that may overcome internet connection limitations [20].

A limitation of this study was the small number of workers included, none of whom required interpreters. Its comprehensibility is therefore unknown among those who require interpreters. Furthermore, these workers were recruited consecutively and recently from the clinic and consequently are not necessarily representative of the wider workforce. Another limitation was that the original instrument was not co-designed with consumers (workers in the SBI), something that we addressed in this study. We were unable to investigate floor or ceiling effects within the scope of this work, which would be a necessary step if the tool is to be used in translated versions.

A strength of the OSEAT is that it has been used and improved for close to 3 years with demonstrably good comprehension by SBI workers.

Future development of the OSEAT will include its translation into other languages using the REDCap multilanguage module.

The commonest languages other than English among the workers attending screening at our center have been Vietnamese, Persian, Chinese, and Arabic, and are therefore a priority for translation and inclusion into the module.

Conclusions

This study has presented the development, comprehension, utility, and refinement of the OSEAT, a purpose-built EDCT for use among SBI workers undergoing assessment for silicosis that included input from workers and has the capacity for modification and use within other silica-exposed occupational settings.

Acknowledgments

This project was funded by WorkSafe Victoria, and FHL was supported by a PhD project stipend from WorkSafe Victoria. The manuscript was prepared by the Monash Centre for Occupational and Environmental Health, Monash University. The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of WorkSafe Victoria.

Data Availability

The datasets generated and analyzed during the current study are not publicly available for ethical and privacy reasons. Patient consent did not include sharing of data external to Monash University. If data need to be shared, additional approval from Monash University HREC and Alfred Hospital HREC would be required.

Authors' Contributions

FHL provided the concept and study design. Input into study design was provided by KWB and DCG. The EDCT was built by CD with input from RH, DG, FHL, MRS, JJM, and KWB. Data acquisition was performed by FHL. Data analysis was completed by FHL, with interpretation by FHL, KWB, and DCG. The manuscript was drafted by FHL, with significant contributions by KWB, RFH, and DCG, and additional editing provided by MRS, JJM, CD, and JF.

Conflicts of Interest

FHL was supported by a PhD project stipend from WorkSafe Victoria. The authors have no further interests to declare.

Multimedia Appendix 1
Occupational history data collection template: stone benchtop industry, and other silica and nonsilica exposed jobs.
[PDF File, 538 KB - humanfactors_v12i1e64111_app1.pdf]

References

1. Barnes H, Glaspole I. Occupational interstitial lung diseases. Immunol Allergy Clin North Am 2023 May;43(2):323-339. [doi: 10.1016/j.iac.2023.01.006] [Medline: 37055091]
2. Austin EK, James C, Tessier J. Early detection methods for silicosis in Australia and internationally: a review of the literature. Int J Environ Res Public Health 2021 Jul 31;18(15):8123. [doi: 10.3390/ijerph18158123] [Medline: 34360414]
3. Hoy RF, Baird T, Hammerschlag G, et al. Artificial stone-associated silicosis: a rapidly emerging occupational lung disease. Occup Environ Med 2018 Jan;75(1):3-5. [doi: 10.1136/oemed-2017-104428] [Medline: 28882991]
4. Shtraichman O, Blanc PD, Ollech JE, et al. Outbreak of autoimmune disease in silicosis linked to artificial stone. Occup Med 2015 Aug;65(6):444-450. [doi: 10.1093/occmed/kqv073] [Medline: 26070814]
5. Leso V, Fontana L, Romano R, Gervetti P, Iavicoli I. Artificial stone associated silicosis: a systematic review. Int J Environ Res Public Health 2019 Feb 16;16(4):568. [doi: 10.3390/ijerph16040568] [Medline: 30781462]
6. Kramer MR, Blanc PD, Fireman E, et al. Artificial stone silicosis [corrected]: disease resurgence among artificial stone workers. Chest 2012 Aug;142(2):419-424. [doi: 10.1378/chest.11-1321] [Medline: 22383661]
7. Grubstein A, Shtraichman O, Fireman E, Bachar GN, Noach-Ophir N, Kramer MR. Radiological evaluation of artificial stone silicosis outbreak: emphasizing findings in lung transplant recipients. J Comput Assist Tomogr 2016;40(6):923-927. [doi: 10.1097/RCT.0000000000000454] [Medline: 27680410]
8. Rosengarten D, Fox BD, Fireman E, et al. Survival following lung transplantation for artificial stone silicosis relative to idiopathic pulmonary fibrosis. Am J Ind Med 2017 Mar;60(3):248-254. [doi: 10.1002/ajim.22687] [Medline: 28145560]
9. Pérez-Alonso A, Córdoba-Doña JA, Millares-Lorenzo JL, Figueroa-Murillo E, García-Vadillo C, Romero-Morillos J. Outbreak of silicosis in Spanish quartz conglomerate workers. Int J Occup Environ Health 2014;20(1):26-32. [doi: 10.1179/2049396713Y.0000000049] [Medline: 24804337]
10. Pascual S, Urrutia I, Ballaz A, Arrizubieta I, Altube L, Salinas C. Prevalence of silicosis in a marble factory after exposure to quartz conglomerates. Arch Bronconeumol 2011 Jan;47(1):50-51. [doi: 10.1016/j.arbres.2010.09.004] [Medline: 21190766]

11. Pascual Del Pobil Y Ferré MA, García Sevila R, García Rodenas MDM, Barroso Medel E, Flores Reos E, Gil Carbonell J. Silicosis: a former occupational disease with new occupational exposure scenarios. Rev Clin Esp (Barc) 2019;219(1):26-29. [doi: 10.1016/j.rce.2018.06.006] [Medline: 30293675]
12. Pérez-Alonso A, Córdoba-Doña JA, León-Jiménez A. Silicosis caused by artificial quartz conglomerates: keys to controlling an emerging disease. Arch Bronconeumol (Eng Ed) 2019 Jul;55(7):394-395. [doi: 10.1016/j.arbr.2019.01.002]
13. Newbigin K, Parsons R, Deller D, Edwards R, McBean R. Stonemasons with silicosis: Preliminary findings and a warning message from Australia. Respirology 2019 Dec;24(12):1220-1221. [doi: 10.1111/resp.13672] [Medline: 31407419]
14. Kumarasamy C, Pisaniello D, Gaskin S, Hall T. What do safety data sheets for artificial stone products tell us about composition? A comparative analysis with physicochemical data. Ann Work Expo Health 2022 Aug 7;66(7):937-945. [doi: 10.1093/annweh/wxac020] [Medline: 35411922]
15. Barnes H, Goh NSL, Leong TL, Hoy R. Silica-associated lung disease: an old-world exposure in modern industries. Respirology 2019 Dec;24(12):1165-1175. [doi: 10.1111/resp.13695] [Medline: 31517432]
16. Hoy RF, Glass DC, Dimitriadis C, Hansen J, Hore-Lacy F, Sim MR. Identification of early-stage silicosis through health screening of stone benchtop industry workers in Victoria, Australia. Occup Environ Med 2021 Apr;78(4):296-302. [doi: 10.1136/oemed-2020-106897] [Medline: 33115923]
17. Ehrlich R, Akugizibwe P, Siegfried N, Rees D. The association between silica exposure, silicosis and tuberculosis: a systematic review and meta-analysis. BMC Public Health 2021 May 20;21(1):953. [doi: 10.1186/s12889-021-10711-1] [Medline: 34016067]
18. Hoy RF, Dimitriadis C, Abramson M, et al. Prevalence and risk factors for silicosis among a large cohort of stone benchtop industry workers. Occup Environ Med 2023 Aug;80(8):439-446. [doi: 10.1136/oemed-2023-108892] [Medline: 37328266]
19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
20. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
21. Glass DC, Dimitriadis C, Hansen J, Hoy RF, Hore-Lacy F, Sim MR. Silica exposure estimates in artificial stone benchtop fabrication and adverse respiratory outcomes. Ann Work Expo Health 2022 Jan 7;66(1):5-13. [doi: 10.1093/annweh/wxab044] [Medline: 35015818]
22. Hore-Lacy F, Hansen J, Dimitriadis C, et al. Predictors of psychological stress in silica-exposed workers in the artificial stone benchtop industry. Respirology 2022 Jun;27(6):455-461. [doi: 10.1111/resp.14257] [Medline: 35421270]
23. Hore-Lacy F, Gwini S, Glass DC, et al. Psychometric properties of the Perceived Stress Scale (PSS-10) in silica-exposed workers from diverse cultural and linguistic backgrounds. BMC Psychiatry 2024 Mar 5;24(1):181. [doi: 10.1186/s12888-024-05613-6] [Medline: 38439053]
24. Health and Safety Executive. Controlling Airborne Contaminants at Work: A Guide to Local Exhaust Ventilation (LEV) - HSG258: Health and Safety Executive; 2017:111 URL: https://www.hse.gov.uk/pubns/books/hsg258.htm [accessed 2025-01-30]
25. Health and Safety Executive. Respiratory Protective Equipment at Work: A Practical Guide - HSG53: Health and Safety Executive; 2013:59 URL: https://www.hse.gov.uk/pubns/books/hsg53.htm [accessed 2025-01-30]
26. WorkSafe Victoria. Managing Exposure to Crystalline Silica: Engineered Stone Compliance Code - WSV1999/02/0822: WorkSafe Victoria; 2022:68 URL: https://www.worksafe.vic.gov.au/resources/archived-compliance-code-managing-exposure-crystalline-silica-engineered-stone [accessed 2025-01-30]
27. Denton E, Hore-Lacy F, Radhakrishna N, et al. Severe Asthma Global Evaluation (SAGE): an electronic platform for severe asthma. J Allergy Clin Immunol Pract 2019;7(5):1440-1449. [doi: 10.1016/j.jaip.2019.02.042] [Medline: 30954467]
28. Dillon DG, Pirie F, Rice S, et al. Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. J Clin Epidemiol 2014 Dec;67(12):1358-1363. [doi: 10.1016/j.jclinepi.2014.06.012] [Medline: 25135245]
29. OME Cleveland Clinic Orthopaedics. Implementing a scientifically valid, cost-effective, and scalable data collection system at point of care: The Cleveland Clinic OME cohort. J Bone Joint Surg Am 2019 Mar 6;101(5):458-464. [doi: 10.2106/JBJS.18.00767] [Medline: 30845040]
30. Brown J, Bhatnagar M, Gordon H, Goodner J, Cobb JP, Lutrick K. An electronic data capture tool for data collection during public health emergencies: development and usability study. JMIR Hum Factors 2022 Jun 9;9(2):e35032. [doi: 10.2196/35032] [Medline: 35679114]
31. Fazio JC, Gandhi SA, Harrison RJ. Workers at risk of silicosis-ongoing overexposure and lack of medical surveillance-reply. JAMA Intern Med 2024 Feb 1;184(2):225-226. [doi: 10.1001/jamainternmed.2023.6626] [Medline: 38109093]
32. Wagner GR, Michaels D. Preventing the continuing tragedy of silicosis. JAMA Intern Med 2024 Feb 1;184(2):223-224. [doi: 10.1001/jamainternmed.2023.6629] [Medline: 38109105]
33. Chen S, Liu M, Xie F. Global and national burden and trends of mortality and disability-adjusted life years for silicosis, from 1990 to 2019: results from the Global Burden of Disease study 2019. BMC Pulm Med 2022 Jun 21;22(1):240. [doi: 10.1186/s12890-022-02040-9] [Medline: 35729551]

34. Wang H, Abbas KM, Abbasifard M, et al. Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct;396(10258):1160-1203. [doi: 10.1016/S0140-6736(20)30977-6] [Medline: 33069325]
35. Hoy RF, Chambers DC. Silica-related diseases in the modern world. Allergy 2020 Nov;75(11):2805-2817. [doi: 10.1111/all.14202] [Medline: 31989662]
36. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-source electronic health record systems for low-resource settings: systematic review. JMIR Med Inform 2017 Nov 13;5(4):e44. [doi: 10.2196/medinform.8131] [Medline: 29133283]
37. Maré IA, Kramer B, Hazelhurst S, et al. Electronic Data Capture System (REDCap) for health care research and training in a resource-constrained environment: technology adoption case study. JMIR Med Inform 2022 Aug 30;10(8):e33402. [doi: 10.2196/33402] [Medline: 36040763]
38. Marroquin Rivera A, Rosas-Romero JC, Castro SM, et al. Implementing a Redcap-based research data collection system for mental health. Rev Colomb Psiquiatr (Engl Ed) 2021 Jul;50 Suppl 1(Suppl 1):110-115. [doi: 10.1016/j.rcpeng.2021.06.004] [Medline: 34257053]

Abbreviations

AS: artificial stone
CNC: Computer Numerical Control
CT: computed tomography
EDCT: electronic data capture tool
OSEAT: Occupational Silica Exposure Assessment Tool
RCS: respirable crystalline silica
REDCap: Research Electronic Data Capture
SBI: stone benchtop industry

Edited by A Kushniruk; submitted 09.07.24; peer-reviewed by C Berriault, RN Carey, SJL Linde; revised version received 14.11.24; accepted 24.11.24; published 24.02.25.

Please cite as:
Hore-Lacy F, Dimitriadis C, Hoy RF, Jimenez-Martin J, Sim MR, Fisher J, Glass DC, Walker-Bone K
Screening Workers for Occupational Exposure to Respirable Crystalline Silica: Development and Usability of an Electronic Data Capture Tool
JMIR Hum Factors 2025;12:e64111
URL: https://humanfactors.jmir.org/2025/1/e64111
doi:10.2196/64111

Collecting Real-World Data via an In-Home Smart Medication Dispenser: Longitudinal Observational Study of Survey Panel Persistency, Response Rates, and Psychometric Properties

Benjamin Ogorek[1], PhD; Thomas Rhoads[1], MBA; Erica Smith[2], PhD

[1]Spencer Health Solutions Inc, 2501 Aerial Center Pkwy, Suite 100, Morrisville, NC, United States

[2]WCG, Princeton, NJ, United States

Corresponding Author:
Benjamin Ogorek, PhD
Spencer Health Solutions Inc, 2501 Aerial Center Pkwy, Suite 100, Morrisville, NC, United States

Abstract

Background: A smart medication dispenser called "spencer" is a novel generator of longitudinal survey data. The patients dispensing medication act as a survey panel and respond to questions about quality of life and patient-reported outcomes.

Objectives: Our goal was to evaluate panel persistency, survey response rates, reliability, and validity of surveys administered via spencer to 4138 polychronic patients residing in the United States and Canada.

Methods: Patients in a Canadian health care provider's program were included if they were dispensing via spencer in the June 2021 to February 2024 time frame and consented to have their data used for research. Panel persistency was estimated via discrete survival methods for 2 years and survey response rates were computed for 1 year. Patients were grouped by mean response rates in the 12th month (<90% vs ≥90%) to observe differential response rate trends. For reliability and validity, we used a spencer question about recent falls with ternary responses value-coded −1, 0, and 1. For reliability, we computed Pearson correlation between mean scores over 2 years of survey responses, and transitions between mean score intervals of [0, 0.5), [−0.5, 0.5), and [0.5, 1]. For validity, we measured the association between the falls question and known factors influencing fall risk: age, biological sex, quality of life, physical and emotional health, and use of selective serotonin reuptake inhibitors or serotonin-norepinephrine reuptake inhibitors, using repeated-measures regression for covariates and Kendall τ for concomitant spencer questions.

Results: From 4138 patients, dispenser persistency was 68.3% (95% CI 66.8% - 69.8%) at 1 year and 51% (95% CI 49% - 53%) at 2 years. Within the cohort observed beyond 1 year, 82.3% (1508/1832) kept surveys enabled through the 12th month with a mean response rate of 84.1% (SD 26.4%). The large SD was apparent in the subgroup analysis, where a responder versus nonresponder dichotomy was observed. For 234 patients with ≥5 fall risk responses in each of the first 2 years, the Pearson correlation estimate between yearly mean scores was 0.723 (95% CI 0.630 - 0.798). For mean score intervals [0, 0.5), [−0.5, 0.5), and [0.5, 1], self-transitions were the most common, with 59.8% (140/234) of patients starting and staying in [0.5, 1]. Fall risk responses were not significantly associated with sex (P=.66) or age (P=.76) but significantly related to selective serotonin reuptake inhibitor or serotonin-norepinephrine reuptake inhibitor usage, quality of life, depressive symptoms, physical health, disability, and trips to the emergency room (P<.001).

Conclusions: A smart medication dispenser, spencer, generated years of longitudinal survey data from patients in their homes. Panel attrition was low, and patients continued to respond at high rates. A fall risk measure derived from the survey data showed evidence of reliability and validity. An alternative to web-based panels, spencer is a promising tool for generating patient real-world data.

(JMIR Hum Factors 2025;12:e60438) doi:10.2196/60438

KEYWORDS

real-world data; real-world evidence; patient-reported outcomes; longitudinal studies; survey methods

Introduction

Background

The use of patient data collected in real-world settings has never been more impactful. The US Food and Drug Administration's Real-World Evidence (RWE) Program has elevated real-world data (RWD) as a tool to support new indications for already approved drugs [1-3], the European Medicines Agency has published their RWE framework [4], and Canada's Drug and Health Technology Agency has published their guidance document [5].

RWD may take the form of claims records, electronic health records, registries, or patient-generated data, with patient-reported outcomes (PROs) as an important subset. Longitudinal surveys, where patients are surveyed at 2 or more points in time, generate data that allow for the analysis of within-unit change as well as aggregations over time [6]. This results in greater "causal leverage" than cross-sectional surveys [7] and is ideal for submissions to regulatory bodies.

Web-based panels, or "registered persons who have agreed to take part in online studies on a regular basis," rose in attractiveness with the proliferation of the web [8]. Recently, however, shortcomings of longitudinal studies based on web-based panels have undermined their reputation as a high-quality data source. Panel attrition, where subjects in earlier waves cease to respond in later waves, has become worse since the 1990s [7,9-11]. While web-based panel data are also prone to quality problems (eg, false answers, careless responses, and multiple panel memberships [12]), these problems have been exacerbated by innovations in automation and improvements in large language models, where human reviewers are unable to consistently detect automated responses [13]. This has become a corrupting force in web-based survey sampling [14].

Amazon Mechanical Turk (MTurk) was considered a representative and convenient source of web-based longitudinal survey data [15] but has seen its reputation deteriorate within the last decade. For example, a study that used MTurk to build a diabetes panel failed after only 5.8% (13/224) were deemed eligible for future survey research [16]. Researchers noted declines in MTurk data quality starting around summer 2018, as evidenced by degraded psychometric properties of well-understood metrics [17]. A warning was issued in the journal *Perspectives on Psychological Science* after an exercise revealed that only 2.6% (14/529) of MTurk samples were valid [18].

Alternatives to web-based panels exist in populations of patients using web-connected hardware, also known as "smart" products. One interesting subset is the population of patients using smart medication dispensers, as these products sit in the home amidst a public health need for digital adherence solutions [19]. A 2023 review of smart medication adherence products reviewed the features of 51 products without mention of survey administration capabilities [20]. One of these products, a dispenser named "spencer" [21,22] (manufactured by Spencer Health Solutions, Inc), has a touch screen display that allows it to administer survey questions following on-time medication dispenses (Figure 1).

At the time of writing, in-home spencer devices have generated more than 3 million longitudinal responses from more than 4000 unique patients to quality of life and PRO measures from a polychronic population residing in the United States and Canada. These are patients of Canadian health care provider Custom Health, Inc, a company offering "a personalized, connected service that goes beyond medication management and ensures medications are working as they should" [23]. Patients or caregivers can express interest directly via a collection of sign-up forms [24,25] or they may be directed to spencer via their health plans that have partnered with Custom Health [26]. When a health plan partners with Custom Health, their services are provided to members "who require a high degree of clinical oversight, those managing multiple medications or those experiencing medication adherence challenges" [27].

Figure 1. Key components of the spencer smart medication dispenser.

Objectives

The study's aim was to evaluate the spencer smart medication dispenser as a longitudinal survey platform for a polychronic patient population. Panel persistency, survey response rates, and measurement reliability and validity were assessed.

Methods

Recruitment

Patients of Custom Health were included if they met the criteria enumerated in Textbox 1.

Textbox 1. Inclusion criteria for patients in the study.

1. The patient entered Custom Health's intake process either by self-selection or based on the recommendation of a health care provider.
2. Custom Health professionals decided to pair the patient with a spencer smart medication dispenser.
3. The patient agreed to the Custom Health consent form.
4. The patient agreed to the Spencer Health Solutions End User License Agreement, permitting his or her deidentified data to be used for research purposes. This occurred on the spencer unit's touch screen.
5. The patient's first scheduled medication dose was between June 3, 2021, and February 14, 2024.
6. The patient dispensed a medication dose by March 14, 2024. In this paper, dispensing medication refers to dispensing multidose packs containing oral solids.

After completing Custom Health's intake process, spencer devices were shipped to patients' homes. Once set up, the devices displayed both current local time and the scheduled time of the next medication dispense via a touch screen display. Refills containing medication strips (multidose adherence packaging) prepared by a pharmacy were shipped to the patients' homes and were inserted by the patient or care nurse into the top of the unit via an electronically controlled door. At scheduled dosing times, the unit alerted the patient through sound, light, and a message on the touch screen display. After the patient pressed the dispense button on the touch screen, the unit dispensed 1 or more medication pouches. After an on-time dispense, a question was presented to those patients who had not explicitly opted out of surveys.

Data Generation and Processing

The question and response mechanism is further elaborated here. If a dose was dispensed on time and the patient had not opted out of surveys, 1 question was displayed on-screen. To answer, a single button press was needed to select from a multiple-choice answer set. This was followed by a review step (also serving as the completeness check) where the patient could confirm the selection or go back and change the answer. If a patient did not make the confirmation in the review step, the questionnaire would not be submitted to the database and later analyzed as an instance of nonresponse. If left attended to, a question would remain on the screen until the next scheduled dose.

In collaboration with health care professionals at Custom Health, 35 survey items were designed to measure the spencer experience, quality of life, and PROs. To avoid copyright infringement, these questions were not taken from any existing validated scale. Questions were scheduled one-to-one with doses in a predefined sequence that repeated indefinitely. Response options were consistently listed from most positive sentiment (eg, "Excellent") toward the top of the screen to least positive sentiment (eg, "Poor") toward the bottom of the screen. Questions were manually answered on test devices in a quality assurance laboratory before being released to patients, and patients could call into a support line to provide feedback regarding the questions or to request that they be turned off.

As is typical of web-based panels, the panel formed by the selection criteria in Textbox 1 constitutes a convenience sample. The target population best described by the sample is polychronic patients taking multiple medications daily. Since the surveys were administered as part of routine patient monitoring, no institutional review board (IRB) approval was needed.

Survey responses were sent to the application database through either cellular connection (the default) or Wi-Fi. In cases where the spencer lost connectivity, a store and forward mechanism sent data to the cloud database once connectivity was reestablished. The database is managed by Spencer Health Solutions that has received both ISO27001 [28] and Data Privacy Framework [29] certifications.

Date of birth, biological sex, and residential postal code were entered into a web-based portal by health care providers when patients were recruited. These fields were retrieved from the application database March 4, 2024. Dates of birth that were within 2 years of the database entry date were replaced with missing values, and age was computed as the difference between the first dispense date and date of birth. Within the United States, 5-digit postal codes were converted into US states via the *zipcodeR* R package. For Canadian postal codes, a function was written that maps the first letter of the postal code to the associated province. Prescription information was created by the pharmacies at the time of refill creation and sent to the database.

Statistical Analysis

Panel Attrition

Patients may leave the spencer dispensing platform for multiple reasons, including life transitions to higher care services or natural death. Leaving the dispensing platform is the primary mechanism of spencer panel attrition. To estimate dispenser persistency, we used the discrete survival analysis framework described by Allison [30], where the periods start on the first

day a patient is scheduled, are 30 days in length, and an attrition event occurs when a patient is not scheduled during an entire 30-day period. There is a *resurrection* mechanism: when a patient is scheduled in a later period after previously meeting the definition for an attrition event, the attrition flag is reset for all previous periods.

For readability, we will refer to a 30-day period as a *month*, 12 thirty-day periods as a *year*, and so on. Furthermore, we will refer to the time in years between the first scheduled dose via spencer and the analysis date as *tenure*. A patient's tenure represents the amount of experience a patient has had with the spencer platform as of an analysis date.

Beyond dispenser attrition, the second source of panel attrition is when patients request that their questions no longer appear on-screen following a dispensing event. To study this phenomenon, we computed rates of requested question discontinuation for the first 12 months of tenure for patients who remained dispenser persistent for more than a year.

When pursuing the subset of patients who were dispenser persistent for more than a year, the subset taken was patients who remained on the spencer platform through the 14th month. In addition to our operational month being shorter on average than a calendar month (by a fraction of a day), the 2 additional months of persistency provided a buffer against the decreased interaction with the device that often precedes full platform discontinuation.

Survey Response Rates

On the survey platform, nonresponse occurs when patients do not enter a response after a question is displayed and the question is cleared. We computed rates of nonresponse by month and plotted the resulting series. We knew from prior analyses that some patients consistently respond to the questions, and we wanted to observe this phenomenon. For patients who were still receiving surveys in the 12th month, we created 2 groups: those with 12th-month response rates of <90% and those with 12th-month response rates of ≥90%. For both groups, we computed the frequency of patients, plotted response rates by month, and provided a qualitative description of the patterns observed.

Psychometric Analysis

Reliability

A reliability analysis in the context of a platform requires a narrowing of focus to a specific measure, as both reliable and unreliable measures may be generated from any platform. Inspired by the Falls Efficacy Scale-International, a reliable measure of fear of falling known to be related to both past and future falls [31,32], we chose an existing question from our rotation that asks the patient about recent falls. Hereafter referred to as Q_FALL, the question text read "Have you experienced a fall in the past month?" The response options were "No," "Not Sure," and "Yes" (a 1-letter variation in capitalization occurring after September 2022, where "Not sure" was replaced with "Not Sure"), which were value-coded as 1, 0, and −1, respectively.

One conceptualization of reliability is test-retest reliability and can be quantified using Pearson correlation between a measure's values at 2 time points [33]. For a comparative baseline in the literature, Falls Efficacy Scale-International measurements taken by the same patients at different time intervals had Pearson correlations ranging from 0.66 to 0.83 for measurements taken up to a year apart [34].

The Pearson correlation coefficient is known to suffer bias when distributional assumptions are violated, a concern because Q_FALL has only 3 response levels and there were different response counts between patients and years. The use of averages and bootstrap resampling were thus employed to address these factors. First, we limited attention to a subset of 234 patients from the persistency analysis who answered Q_FALL at least 5 times in both a full first year and a second year of tenure, hereafter referred to as year 1 and year 2. Second, we used the bootstrap to obtain a bias-corrected estimate of the Pearson correlation along with a nonparametric 95% CI [35]. This allowed us to perform inference on the coefficient of determination (R^2) for the equivalent regression of the year 2 means regressed on the year 1 means.

Averaging the ternary scores allowed us to work on a continuum where rare fallers and never fallers appear close together on the resulting scale, a notion supported by similarities between these groups in a 1-year cohort study [36]. To circumvent the limitations of a linear correlation analysis, we performed an additional discrete state transition analysis. We examined the frequency of transitions to and from mean Q_FALL scores of [−1, −0.5), [−0.5, 0.5), and [0.5, 1.0] in year 1 and year 2, expecting self-transitions to be the most frequent.

Validity

To assess the convergent validity of the recent falls question administered via spencer, the mean scores for year 1 and year 2 were compared with the following established risk factors of fall risk: increased age, biological sex, previous fall frequency, low quality of life, depressive symptoms, physical impairment, and medication use [31,34,37]. Many patients in this population were prescribed selective serotonin reuptake inhibitors (SSRIs) or serotonin-norepinephrine reuptake inhibitors (SNRIs), and these are associated with falls in the older adults [38,39]. A meta-analysis found that 95% (70/74) of studies reported gender or sex differences in fall-related outcomes with females at a higher risk than males [40]. Canonically, increased age is a risk factor for falls [41]. The validity analysis was split into 2 parts, each based on the 234-patient subset from the reliability analysis.

We first conducted an analysis of the relationship between raw Q_FALL values and covariates age, sex, and SSRI or SNRI usage, as these were known before any responses were received (medication can be discontinued but medication classes tend to be stable within patient). To accommodate the repeated measures received from each patient, we used a generalized estimating equation approach to model the relationship between the coded value of Q_FALL and an exchangeable working correlation structure. This was accomplished with the *geepack* package in R, which reports SEs that are robust to both the choice of working correlation structure and nonnormality of the response. For age and SSRI or SNRI usage, we expected to see negative

relationships. For biological sex, we expected that female patients would be associated with lower mean Q_FALL than male patients.

For evidence of association between Q_FALL and other relevant variables, including quality of life, depressive symptoms, and hospital visits, we selected the questions listed in Table 1 as contemporaneous survey-based measures that had face validity for concepts of interest. Their responses are integer-coded and arranged by sentiment, and thus we expected positive correlations with Q_FALL.

The robust Kendall τ measure was used to test for associations, as the sample sizes of the questions from Table 1 may be arbitrarily small within patient. Kendall τ is more appropriate for ties and has an accompanying 2-sided nonparametric test for testing the null hypothesis of zero association [35]. For a nonparametric 95% CI on τ, we used the *kendall.ci* function from the R package *NSM3* [42], which provides a bootstrap CI.

Table . Standard spencer questions relating to known risk factors of falling.

Question text	Possible responses	Values coded	Construct[a]
Rate your recent quality of life.	Excellent | Very good | Good | Fair | Poor	5 | 4 | 3 | 2 | 1	Quality of life
How is your emotional health today?[b]	Excellent | Very good | Good | Fair | Poor	5 | 4 | 3 | 2 | 1	Depression
How would you rate your physical health today?	Excellent | Very good | Good | Fair | Poor	5 | 4 | 3 | 2 | 1	Physical health
Rate your ability to perform activities today.	Excellent | Very good | Good | Fair | Poor	5 | 4 | 3 | 2 | 1	Ability or disability
Are you able to accomplish what you have planned today?	Completely | Mostly | Moderately | A little | Not at all	5 | 4 | 3 | 2 | 1	Ability or disability
Hospital, ER[c], or urgent care in the past month?	No | Not sure | Yes	1 | 0 | −1	ER visits from falls

[a]Construct is based on face validity of the spencer standard questions.

[b]This question has been in rotation for multiple years, but in September 2022, the number of responses changed from 3 (“Poor,” “Good,” and “Excellent”) to 5 (“Poor,” “Fair,” “Good,” “Very good,” and “Excellent”). We coded the 3-response set as 1, 3, and 5, and the 5-response set as 1, 2, 3, 4, and 5, respectively.

[c]ER: emergency room.

Ethical Considerations

This study used operational data collected from a commercial medication dispensing system used in routine patient care and was not subject to IRB review requirements, so IRB approval was not pursued. Users of the spencer device provided consent for data collection through the End User License Agreement, which covers the collection of medication adherence data and responses to quality of life and PRO surveys as part of the system’s standard operation. No additional compensation was provided to users beyond the normal terms of their device usage agreement. All data analyzed in this study were deidentified prior to analysis. Spencer Health Solutions has achieved both ISO27001 [28] and Data Privacy Framework [29] certifications, and the system uses industry-standard encryption and security measures.

This research analyzed data collected during standard clinical care and device usage. All results are presented as anonymous aggregate statistics. The original data collection occurred as part of routine clinical practice, with patients providing consent for research use through the device terms of service and care management agreement. Under Canadian TCPS 2 Article 2.4, research ethics board review is not required for research that relies exclusively on secondary use of anonymous information where the process does not generate identifiable information. Under US regulation 45 CFR 46.104(d)(4)(ii), IRB review is not required when information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects, the investigator does not contact the subjects, and the investigator will not reidentify subjects.

Results

Patient Population

The patient population was majority female (2552/4133, 61.7%), with 0.1% (5/4138) of the biological sex fields missing. The mean age was 54.4 years (SD 19.9, range 5-104 years). Most patients (3736/4138, 90.3%) resided in Canada, with 1 address unmapped at the country level. Patients were scheduled to take multiple drugs per day (mean 9.6, SD 5.1). Of the 2805 unique compounds scheduled during the observation window, 70.2% (1970/2805) were mapped to an Anatomical Therapeutic Chemical classification system second-level code, with a modified “Vitamins & Supplements” that included dietary supplements. Table 2 contains the 20 most frequently observed subgroups observed during the observation window.

Table . Patient demographics, geographic distribution, and medication usage among the 4138 patients studied.

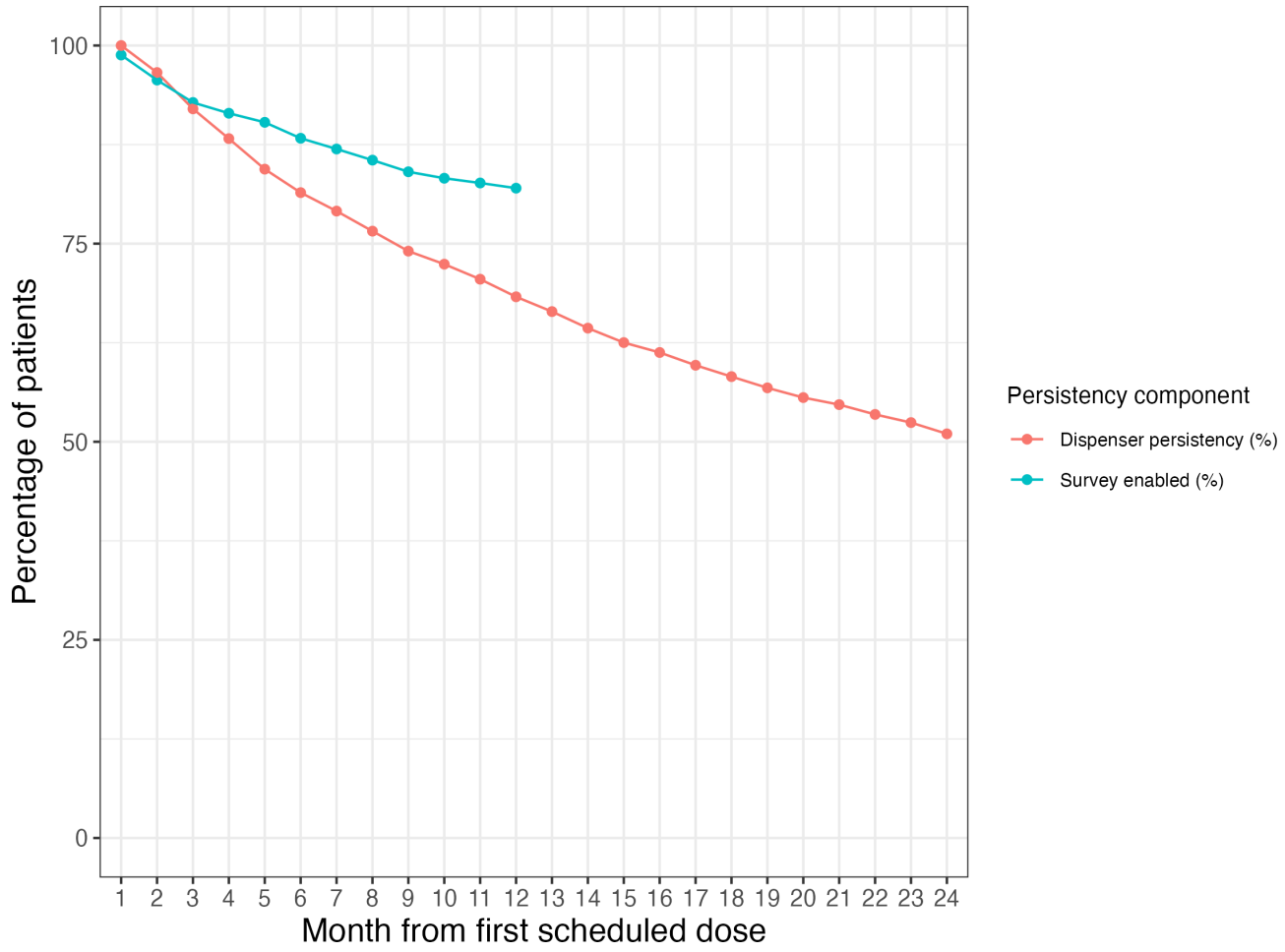
Section and variable	Patients (N=4138)
Patient demographics, n (%)	
Sex	
Female	2552 (61.7)
Male	1581 (38.2)
Missing	5 (0.1)
Age (years)	
Valid records, n (%)	4123 (99.6)
Invalid records, n (%)	15 (0.4)
Mean (SD)	54.4 (19.9)
IQR	39 - 70
Age range	5 - 104
Geographic distribution, n (%)	
Country	
Canada	3736 (90.3)
United States	401 (9.7)
Missing or other	1 (0)
Canadian provinces	
Ontario	2083 (50.3)
British Columbia	1149 (27.8)
Saskatchewan	485 (11.7)
Other	19 (0.5)
US states	
Tennessee	221 (5.3)
Missouri	117 (2.8)
California	32 (0.8)
Ohio	24 (0.6)
Other	7 (0.2)
Medication usage (Anatomic Therapeutic Chemical codes, second level), n (%)	
Psychoanaleptics	3040 (73.5)
Vitamins & supplements	1917 (46.3)
Lipid-modifying agents	1848 (44.7)
Drugs for acid-related disorders	1788 (43.2)
Antiepileptics	1717 (41.5)
Agents acting on the renin-angiotensin system	1664 (40.2)
Psycholeptics	1503 (36.3)
Drugs used in diabetes	1317 (31.8)
Beta-blocking agents	1107 (26.8)
Antithrombotic agents	1050 (25.4)
Calcium channel blockers	893 (21.6)
Diuretics	822 (19.9)
Thyroid therapy	741 (17.9)

Section and variable	Patients (N=4138)
Urologicals	592 (14.3)
Analgesics	543 (13.1)
Anti-inflammatory and antirheumatic products	378 (9.1)
Antihypertensives	345 (8.3)
Anti-Parkinson drugs	309 (7.5)
Drugs for constipation	295 (7.1)
Antihistamines for systemic use	289 (7)

Panel Attrition

From 4138 patients, dispensing persistency was estimated to be 68.3% (95% CI 66.8% - 69.8%) at year 1 and 51% (95% CI 49% - 53%) at year 2. Among the patients who stayed on the dispensing platform past year 1, 82.3% (1508/1832) kept surveys enabled through the 12th month. The rates of question opt-out slowed during the year, as can be seen in Figure 2.

Figure 2. Dispenser persistency and percentage of patients with surveys enabled by month on platform.



Survey Response Rates

Among the 1508 patients who kept their surveys enabled through year 1, the mean response rate was 95.6% (SD 11.9%) in the first month and 84.1% (SD 26.4%) in the 12th month, with the rate of decline slowing in the second half of the year (Figure 3). For patients with surveys enabled in the 12th month, 67.9% (1024/1508) had response rates at or above 90% and 32.1% (484/1508) had response rates below 90%. Figure 4 shows the trajectories of both groups, where the high-response group maintained near perfect response rates while the low-response group experienced a substantial decline by the 12th month.

Figure 3. Mean survey response rate (%) for patients with enabled surveys by month on platform.

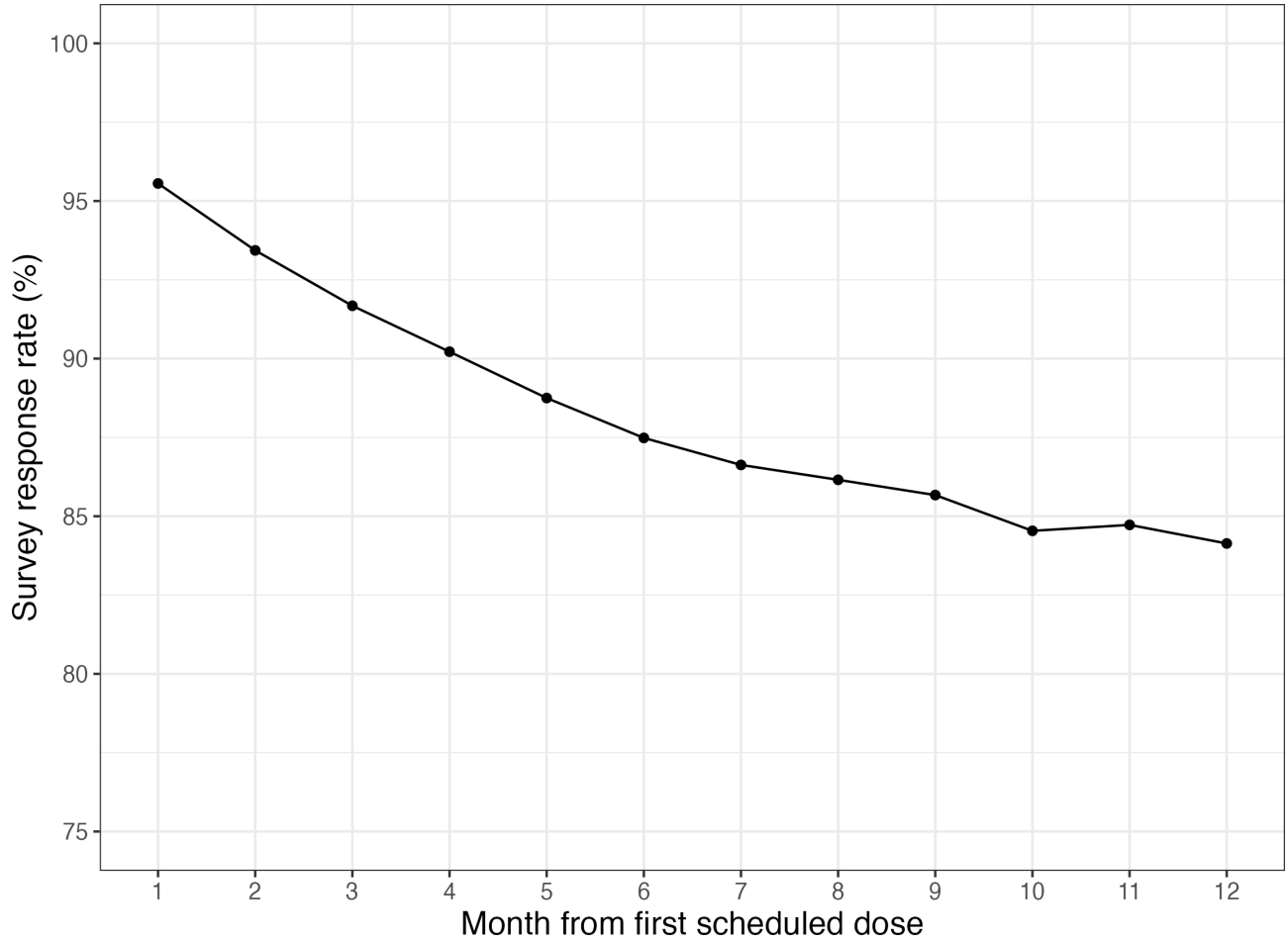
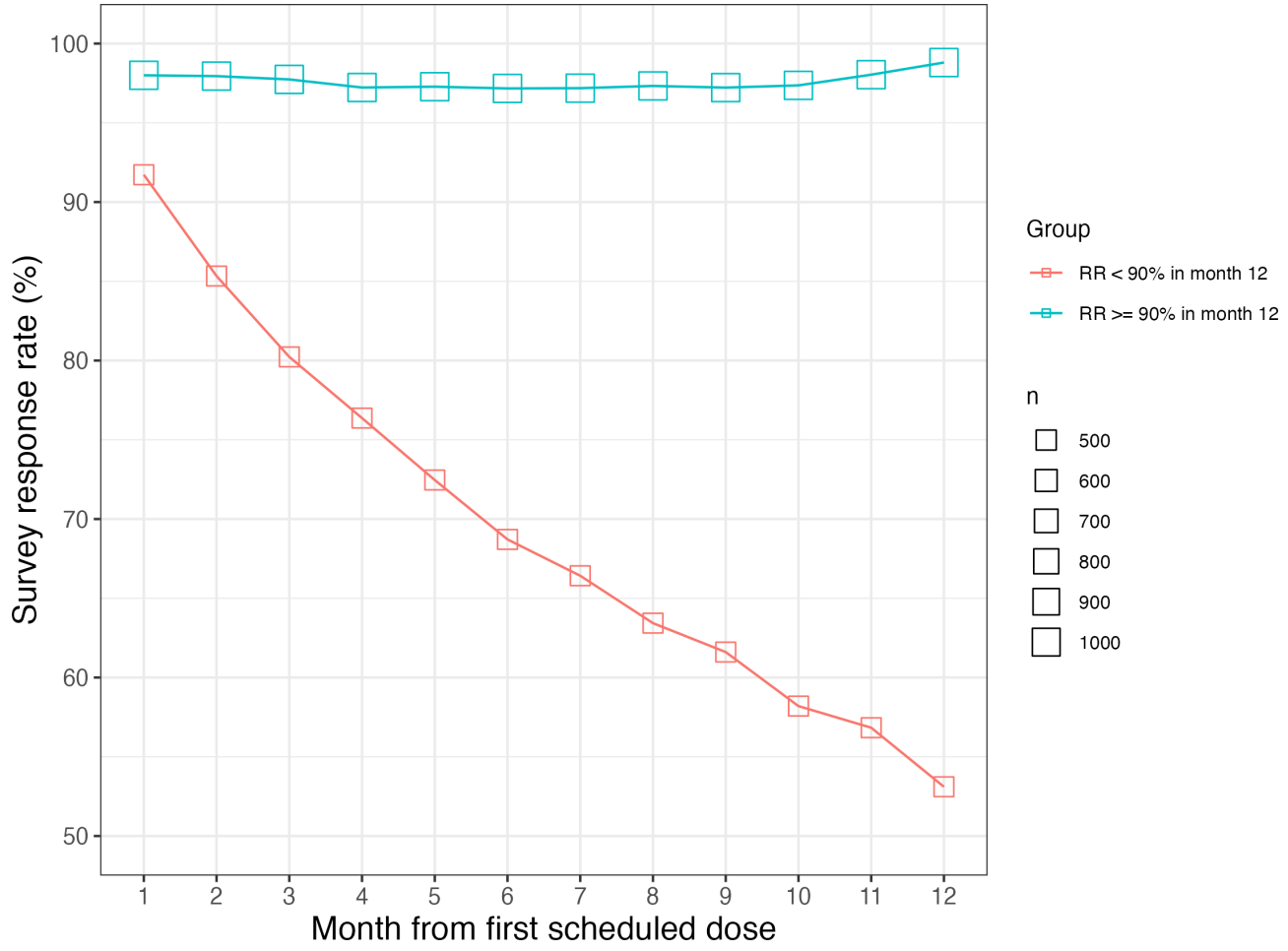


Figure 4. Mean survey RR by month on platform, grouped by RR in the 12th month. RR: response rate.



Reliability

Among 234 patients, the bootstrap estimate of the Pearson correlation between the year 1 and year 2 mean Q_FALL was 0.723 (95% CI 0.630 - 0.798), and the estimate of R^2 for the equivalent regression on year 2 versus year 1 means was 0.523 (95% CI 0.397 - 0.637).

Mean Q_FALL interval transitions from year 1 to year 2 are shown in Table 3. As hypothesized, self-transitions were the most common, with 59.8% (140/234) of patients starting and staying in [0.5, 1]. These rare fallers in the [0.5, 1.0] interval in year 1 remained in this interval during year 2 in 83.8% (140/167) of cases. For the frequent fallers in the [−1.0, −0.5) interval during year 1, 66.7% (12/18) remained during year 2.

Table . Mean score range transitions from year 1 to year 2[a].

Year 1 and year 2 (n=234)		Frequency, n (%)
Frequent fallers [−1.0, −0.5) (n=18)		
	[−1.0, −0.5)	12 (66.7)
	[−0.5, 0.5)	5 (27.8)
	[0.5 to 1.0]	1 (5.6)
Occasional fallers [−0.5, 0.5) (n=49)		
	[−1.0, −0.5)	5 (10.2)
	[−0.5, 0.5)	24 (49)
	[0.5 to 1.0]	20 (48.8)
Rare fallers [0.5, 1.0] (n=167)		
	[−1.0, −0.5)	1 (0.6)
	[−0.5, 0.5)	26 (15.6)
	[0.5 to 1.0]	140 (83.8)

[a]A mean score of +1.0 would be a “perfect score” of no reported falls, a mean score of −1.0 indicates that all responses indicated that a recent fall had occurred, and scores in between span the interval (−1.0, 1.0). One patient who moved from [−1.0, −0.5) to [0.5, 1.0] had a score of exactly 0.5 (the boundary), with only 6 measurements in year 2.

Validity

Covariates

For 232 patients, the generalized estimating equation model results of Q_FALL on these covariates are shown in Table 4. The coefficients associating biological sex and age to Q_FALL were not significantly different from zero; this was unexpected. However, the coefficient indicating the presence of an SSRI or SNRI medication was negative and highly significant, indicating more falls in the SSRI or SNRI group adjusted for sex and age.

Table . Generalized estimating equation linear model summary for sex, age, and whether the patient was prescribed a selective serotonin reuptake inhibitor or serotonin-norepinephrine reuptake inhibitor during the observation window.

Coefficient	Estimate	SE	Wald	*P* value
Intercept	0.684	0.127	28.857	<.001[a]
Sex (male)	−0.030	0.069	0.195	.66[b]
Patient age	0.001	0.002	0.092	.76[b]
SSRI[c] or SNRI[d]	−0.232	0.062	14.092	<.001[b]

[a]The *P* value corresponding to the hypothesis of the intercept being zero is included by convention but is not a meaningful statistic.

[b]These *P* values correspond to the 2-sided test of the hypothesis of a zero regression coefficient.

[c]SSRI: selective serotonin reuptake inhibitor.

[d]SNRI: serotonin-norepinephrine reuptake inhibitor.

Contemporaneous Outcomes From the Spencer

The analysis of correlations between Q_FALL and other contemporaneous spencer questions, based on Kendall τ, is shown in Table 5. Interpretation of correlation coefficients varies, for example, 0.2 may be characterized as either “weak” or “poor,” and 0.3 as “weak,” “moderate,” or “fair” [43], and in the bivariate normal case, a τ value of 0.200 corresponds to a Pearson correlation of 0.309 [44]. While the strength of association between mean Q_FALL and the contemporaneous response outcomes was consistently weak to moderate, *P* values were uniformly small, indicating positive relationships of these questions with the measure of recent falls.

Table . Contemporaneous association between survey questions administered via spencer.

Question	Patients, n[a]	Average responses per patient, n	Kendall τ	95% bootstrap CI for τ	*P* value[b]
Rate your recent quality of life.	233	14	0.15	0.058-0.244	<.001
How is your emotional health today?	233	54	0.21	0.120-0.293	<.001
How would you rate your physical health today?	164	8	0.18	0.071-0.296	<.001
Rate your ability to perform activities today.	232	35	0.23	0.140-0.316	<.001
Are you able to accomplish what you have planned today?	197	9	0.18	0.083-0.276	<.001
Hospital, ER[c], or urgent care in the past month?	192	9	0.20	0.096-0.299	<.001

[a]Number of unique patients who responded to each question at least once and also responded at least 5 times to Q_FALL in year 1 and year 2.

[b]Derived from Kendall τ test, a nonparametric hypothesis test used to measure the ordinal association between 2 variables.

[c]ER: emergency room.

Discussion

Principal Findings

Although its primary function is dispensing medication, the spencer platform doubled as a web-based longitudinal panel where polychronic patients answered survey questions at high rates and exhibited low panel attrition over years of platform tenure. Measures generated from the responses where stable through time (ie, evidence of reliability) and were associated with other theoretically related variables (ie, evidence of validity). For polychronic patients residing in the US and Canada, the home medication dispenser is a promising source of reliable and valid measures of important health constructs.

As with all survey panels, there was attrition and nonresponse. Panel attrition could be decomposed into attrition from the dispensing platform and survey opt-outs for patients remaining on the dispensing platform. These losses were cumulative. Based on the estimates presented, starting with 100 patients, 68 would still be dispensing via spencer by the end of the first year, with 56 still receiving questions following their dispenses.

In our literature review, persistency was often a serious issue in the context of longitudinal patient studies. For 8 remote digital studies conducted between 2014 and 2019, researchers found that more than half of all participants discontinued their participation within the first week of the study [9]. In a web-based study during the COVID-19 pandemic, of 2734 participants who completed wave 1, only 964 participated in wave 3 [10]. In a study of smartphone app usage to improve oral anticoagulation adherence, a retention rate of 27% at 6 months was reported [11]. Considering these results, keeping more than half of the initial patients actively participating in surveys at the end of the first year represents favorable retention.

The rate of new survey opt-outs also decreased substantially through the year, setting up milder losses in year 2.

By the end of first year, the average survey response rate for patients taking surveys on spencer was 84%. While 80% has been considered excellent in the context of primary care research studies [45], multi-item surveys administered at a single point in time are an imperfect benchmark. Ecological momentary assessment, a survey methodology that addresses phenomena as they occur, typically sees compliance rates from 50% to 90% [46]. By either standard, the response rates observed in spencer surveys were good.

We can speculate on why some patients responded to fewer spencer questions over time than others. Survey fatigue is a well-known phenomenon that occurs when respondents become weary of repeated survey tasks [47], and although surveys administered via spencer are brief, they are frequent. In addition to fatigue, some patients may not have been aware of how the questions were being used to monitor their well-being. Developing interventions to improve survey response rates is a topic for future research.

Noncoverage in web-based surveys, often defined as lack of access to the web, is thought to be a more serious problem than nonresponse, which is an unwillingness to participate [48]. Since every participant has a connection to the web through the device itself (the spencer units have both cell and Wi-Fi connections), there is no noncoverage in the sense of lack of web access, although machines do go offline for varying durations.

With sufficiently high panel persistency and response rates, attention focuses on the quality of the data that are generated. We showed that a measure about recent falls generated from a spencer question exhibited temporal stability, a form of reliability. The measure showed expected associations with most theoretically related variables. One exception was the

demographic factors of age and biological sex, which failed to achieve significance in a regression where medication use was significant ($P<.001$). While additional data may reveal the expected relationships, we surmise that in a polychronic population taking many medications, demographic factors may be weaker predictors of falling than in the general population.

Limitations

First, since the patients studied in this paper were polychronic patients enrolled in a care management program and residing in the United States and Canada, inferences to other populations may not be warranted.

Second, our validity analyses were limited to data collected entirely within the spencer ecosystem. While correlations between spencer survey responses suggest meaningful patterns, these questions, although having face validity, lack validation against established instruments. The observed correlations might partly reflect the consistent presentation format on the device, where responses are always ordered from most to least positive sentiment. Our validity arguments would be substantially stronger with independent measurements, particularly comparisons between spencer responses and validated traditional instruments measuring the same constructs.

Third, this study did not consider sensitivity to change, which is important in the context of RWE because it allows researchers and clinicians to detect change resulting from a minimal intervention [49]. While we focused on reliability by treating fall risk as a stable construct, and although fall risk is sufficiently stable to support our reliability analysis, treating it as static was a limitation of this research. Future research could explore methods for estimating changing states from longitudinal survey data, building on established approaches in the literature [50,51].

Conclusions

Administering longitudinal surveys via spencer, a smart medication dispenser, effectively generated high-quality RWD from patients in their homes. Patients persisted on the platform for years and maintained high response rates. A measure derived from longitudinal surveys assessing fall risk demonstrated both reliability and validity. The performance of spencer as a longitudinal survey platform offers a promising alternative at a time when web-based survey data quality is deteriorating.

Because medication dispensing is a fundamental component of the survey-generating mechanism, RWD from the spencer platform offers an ideal opportunity to study medication effectiveness and health outcomes, providing evidence to support new drug indications and demonstrate relationships between health outcomes and economic factors.

Acknowledgments

This work was funded by Spencer Health Solutions, Inc, a private corporation headquartered in Morrisville, North Carolina. The authors wish to thank University of Maryland professor Greg Hancock for his perspective on presenting this new source of patient-reported outcomes. The authors thank Devi Kodavanti, a member of the data team at Spencer Health Solutions, for computing many reliability coefficients in the past that gave them confidence to do this research. Their resident designer Lew Wilson created the table of contents image. Finally, the authors thank Alan Menius, the scientist behind the evolution of the spencer standard questions.

Data Availability

The dataset analyzed during this study is not publicly available to protect patient privacy and comply with health care privacy regulations, since it contains patient-reported outcomes collected through medication-dispensing devices. While the data are deidentified, they represent longitudinal health-related information of single patients alongside demographic information; it is difficult to guarantee that reidentification, however unlikely, is impossible. However, controlled access to specific portions of the data are available from the corresponding author on reasonable request, which should include (1) a brief research proposal outlining their intended use of the data, and (2) documentation of their institution's data security protocols.

Authors' Contributions

All authors participated in the conceptualization, review, and revision of the final manuscript draft for submission. BO proposed the methodology, performed the formal analysis, and wrote the original draft. TR supervised the research through all stages and performed validation activities.

Conflicts of Interest

Authors BO and TR are current employees of Spencer Health Solutions, Inc, the developer of the spencer platform. Author ES was employed at Spencer Health Solutions during the initial research and manuscript development phase.

References

1. Framework for FDA’s real‐world evidence program. U.S. Food and Drug Administration. 2018. URL: https://www.fda.gov/media/120060/download [accessed 2025-01-23]

2. Klonoff DC. The New FDA Real-World Evidence Program to support development of drugs and biologics. J Diabetes Sci Technol 2020 Mar;14(2):345-349. [doi: 10.1177/1932296819832661] [Medline: 30862182]
3. Considerations for the use of real-world data and real-world evidence to support regulatory decision-making for drug and biological products. U.S. Food and Drug Administration. 2023. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug [accessed 2025-01-23]
4. Real-world evidence framework to support EU regulatory decision-making: report on the experience gained with regulator-led studies from september 2021 to february 2023. European Medicines Agency. 2023. URL: https://www.ema.europa.eu/system/files/documents/report/real-world-evidence-framework-support-eu-regulatory-decision-making-report-experience-gained_en.pdf [accessed 2025-01-23]
5. Guidance for reporting real-world evidence. Canada's Drug Agency. 2023. URL: https://www.cadth.ca/sites/default/files/RWE/MG0020/MG0020-RWE-Guidance-Report-Secured.pdf [accessed 2025-01-23]
6. Lynn P. Methodology of Longitudinal Surveys: John Wiley & Sons, Ltd; 2009. [doi: 10.1002/9780470743874]
7. Hillygus DS, Snell S. In: Atkeson LR, Alvarez RM, editors. Longitudinal Surveys: Issues and Opportunities: Oxford University Press; 2015. [doi: 10.1093/oxfordhb/9780190213299.013.7]
8. Göritz AS, Reinhold N, Batinic B. Online panels. In: Online Social Sciences: Hogrefe & Huber Publishers; 2002:27-47.
9. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
10. Yu T, Chen J, Gu NY, Hay JW, Gong CL. Predicting panel attrition in longitudinal HRQoL surveys during the COVID-19 pandemic in the US. Health Qual Life Outcomes 2022 Jul 6;20(1):104. [doi: 10.1186/s12955-022-02015-8] [Medline: 35794553]
11. Senoo K, Miki T, Ohkura T, et al. A smartphone app to improve oral anticoagulation adherence in patients with atrial fibrillation: prospective observational study. JMIR Mhealth Uhealth 2022 Jan 7;10(1):e30807. [doi: 10.2196/30807] [Medline: 34894626]
12. Hays RD, Liu H, Kapteyn A. Use of Internet panels to conduct surveys. Behav Res Methods 2015 Sep;47(3):685-690. [doi: 10.3758/s13428-015-0617-9] [Medline: 26170052]
13. Lebrun B, Temtsin S, Vonasch A, Bartneck C. Detecting the corruption of online questionnaires by artificial intelligence. Front Robot AI 2023;10:1277635. [doi: 10.3389/frobt.2023.1277635] [Medline: 38371744]
14. Research in the crowdsourcing age, a case study. Pew Research Center. 2016. URL: https://www.pewresearch.org/internet/2016/07/11/research-in-the-crowdsourcing-age-a-case-study/ [accessed 2025-01-23]
15. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
16. Lin YK, Newman S, Piette J. Response consistency of crowdsourced web-based surveys on type 1 diabetes. J Med Internet Res 2023 Aug 18;25:e43593. [doi: 10.2196/43593] [Medline: 37594797]
17. Chmielewski M, Kucker SC. An MTurk crisis? Shifts in data quality and the impact on study results. Soc Psychol Personal Sci 2020 May;11(4):464-473. [doi: 10.1177/1948550619875149]
18. Webb MA, Tangney JP. Too good to be true: bots and bad data from Mechanical Turk. Perspect Psychol Sci 2022 Nov:174569162211200. [doi: 10.1177/17456916221120002] [Medline: 36343213]
19. Hafez G, Aarnio E, Mucherino S, et al. Barriers and unmet educational needs regarding implementation of medication adherence management across Europe: insights from COST Action ENABLE. J Gen Intern Med 2024 Jun;39(15):2917-2926. [doi: 10.1007/s11606-024-08851-2] [Medline: 38941058]
20. Faisal S, Samoth D, Aslam Y, et al. Key features of smart medication adherence products: updated scoping review. JMIR Aging 2023 Dec 19;6:e50990. [doi: 10.2196/50990] [Medline: 38113067]
21. Patel T, Ivo J, Pitre T, Faisal S, Antunes K, Oda K. An in-home medication dispensing system to support medication adherence for patients with chronic conditions in the community setting: prospective observational pilot study. JMIR Form Res 2022 May 19;6(5):e34906. [doi: 10.2196/34906] [Medline: 35587371]
22. Baby B. User Experience with Medication Adherence Technology: Determining Usability by Capabilities [Graduate thesis]. : University of Waterloo; 2024 URL: http://hdl.handle.net/10012/20613 [accessed 2025-01-23]
23. Personal solutions. Custom Health. URL: https://customhealth.com/solutions/personal.html [accessed 2025-01-23]
24. Sign up. Custom Health. URL: https://customhealth.com/signup/ [accessed 2025-01-23]
25. i'm spencer: the future of connected care!. Custom Health. URL: https://customhealth.com/spencer.html [accessed 2025-01-23]
26. Professional solutions. Custom Health. URL: https://customhealth.com/solutions/professional-solutions.html [accessed 2025-01-23]
27. Custom health to provide services to health plan of San Mateo members. PR Newswire. 2022 Nov 29. URL: https://www.prnewswire.com/news-releases/custom-health-to-provide-services-to-health-plan-of-san-mateo-members-301689232.html [accessed 2025-01-23]

28. ISO/IEC 27001:2022: information security, cybersecurity and privacy protection — information security management systems — requirements. International Organization for Standardization. 2022. URL: https://www.iso.org/standard/27001 [accessed 2025-01-23]
29. Data Privacy Framework Program. URL: https://www.dataprivacyframework.gov/ [accessed 2025-01-23]
30. Allison PD. Discrete-time methods for the analysis of event histories. Sociol Methodol 1982;13:61. [doi: 10.2307/270718]
31. Yardley L, Beyer N, Hauer K, Kempen G, Piot-Ziegler C, Todd C. Development and initial validation of the Falls Efficacy Scale-International (FES-I). Age Ageing 2005 Nov;34(6):614-619. [doi: 10.1093/ageing/afi196] [Medline: 16267188]
32. Chen Y, Du H, Song M, et al. Relationship between fear of falling and fall risk among older patients with stroke: a structural equation modeling. BMC Geriatr 2023 Oct 11;23(1):647. [doi: 10.1186/s12877-023-04298-y] [Medline: 37821821]
33. Jhangiani RS, Chiang IC, Cuttler C, Leighton DC. Research Methods in Psychology, 4th edition: Kwantlen Polytechnic University; 2019.
34. Delbaere K, Close JCT, Mikolaizak AS, Sachdev PS, Brodaty H, Lord SR. The Falls Efficacy Scale International (FES-I). A comprehensive longitudinal validation study. Age Ageing 2010 Mar;39(2):210-216. [doi: 10.1093/ageing/afp225] [Medline: 20061508]
35. Bishara AJ, Hittner JB. Reducing bias and error in the correlation coefficient due to nonnormality. Educ Psychol Meas 2015 Oct;75(5):785-804. [doi: 10.1177/0013164414557639] [Medline: 29795841]
36. Asai T, Oshima K, Fukumoto Y, Yonezawa Y, Matsuo A, Misu S. The association between fear of falling and occurrence of falls: a one-year cohort study. BMC Geriatr 2022 May 5;22(1):393. [doi: 10.1186/s12877-022-03018-2] [Medline: 35509040]
37. Dewan N, MacDermid JC. Fall Efficacy Scale-International (FES-I). J Physiother 2014 Mar;60(1):60. [doi: 10.1016/j.jphys.2013.12.014] [Medline: 24856947]
38. Darowski A, Chambers SA, Chambers DJ. Antidepressants and falls in the elderly. Drugs Aging 2009;26(5):381-394. [doi: 10.2165/00002512-200926050-00002] [Medline: 19552490]
39. Park H, Satoh H, Miki A, Urushihara H, Sawada Y. Medications associated with falls in older people: systematic review of publications from a recent 5-year period. Eur J Clin Pharmacol 2015 Dec;71(12):1429-1440. [doi: 10.1007/s00228-015-1955-3] [Medline: 26407688]
40. Sebastiani C, Wong JYX, Litt A, et al. Mapping sex and gender differences in falls among older adults: a scoping review. J Am Geriatr Soc 2024 Mar;72(3):903-915. [doi: 10.1111/jgs.18730] [Medline: 38147460]
41. Ambrose AF, Paul G, Hausdorff JM. Risk factors for falls among older adults: a review of the literature. Maturitas 2013 May;75(1):51-61. [doi: 10.1016/j.maturitas.2013.02.009] [Medline: 23523272]
42. Schneider G, Chicken E, Becvarik R. NSM3: functions and datasets to accompany Hollander, Wolfe, and Chicken—Nonparametric Statistical Methods, Third Edition. The Comprehensive R Archive Network. 2023. URL: https://CRAN.R-project.org/package=NSM3 [accessed 2025-01-23]
43. Akoglu H. User's guide to correlation coefficients. Turk J Emerg Med 2018 Sep;18(3):91-93. [doi: 10.1016/j.tjem.2018.08.001] [Medline: 30191186]
44. Gilpin AR. Table for conversion of Kendall's tau to Spearman's rho within the context of measures of magnitude of effect for meta-analysis. Educ Psychol Meas 1993 Mar;53(1):87-92. [doi: 10.1177/0013164493053001007]
45. Booker QS, Austin JD, Balasubramanian BA. Survey strategies to increase participant response rates in primary care research studies. Fam Pract 2021 Sep 25;38(5):699-702. [doi: 10.1093/fampra/cmab070] [Medline: 34213547]
46. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
47. de Koning R, Egiz A, Kotecha J, et al. Survey fatigue during the COVID-19 pandemic: an analysis of neurosurgery survey response rates. Front Surg 2021 Aug 12;8:690680. [doi: 10.3389/fsurg.2021.690680] [Medline: 34458314]
48. Couper MP, Kapteyn A, Schonlau M, Winter J. Noncoverage and nonresponse in an internet survey. Soc Sci Res 2007 Mar;36(1):131-148. [doi: 10.1016/j.ssresearch.2005.10.002]
49. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. Qual Life Res 1992 Feb;1(1):73-75. [doi: 10.1007/BF00435438] [Medline: 1301117]
50. Le H, Schmidt FL, Putka DJ. The multifaceted nature of measurement artifacts and its implications for estimating construct-level relationships. Organ Res Methods 2009 Jan;12(1):165-200. [doi: 10.1177/1094428107302900]
51. Ozkok O, Zyphur MJ, Barsky AP, Theilacker M, Donnellan MB, Oswald FL. Modeling measurement as a sequential process: Autoregressive Confirmatory Factor Analysis (AR-CFA). Front Psychol 2019;10:2108. [doi: 10.3389/fpsyg.2019.02108] [Medline: 31616338]

Abbreviations

IRB: institutional review board
MTurk: Amazon Mechanical Turk
PRO: patient-reported outcome
Q_FALL: question about recent falls asked on spencer

RWD: real-world data
RWE: real-world evidence
SNRI: serotonin-norepinephrine reuptake inhibitor
SSRI: selective serotonin reuptake inhibitor

Edited by K Adapa; submitted 10.05.24; peer-reviewed by J Misko, LB Iantovics; revised version received 06.11.24; accepted 03.12.24; published 03.02.25.

Please cite as:
Ogorek B, Rhoads T, Smith E
Collecting Real-World Data via an In-Home Smart Medication Dispenser: Longitudinal Observational Study of Survey Panel Persistency, Response Rates, and Psychometric Properties
JMIR Hum Factors 2025;12:e60438
URL: https://humanfactors.jmir.org/2025/1/e60438
doi:10.2196/60438

Comparisons of Physicians', Nurses', and Social Welfare Professionals' Experiences With Participation in Information System Development: Cross-Sectional Survey Study

Susanna Martikainen[1], PhD; Johanna Viitanen[2], DSc; Samuel Salovaara[3], PhD; Ulla-Mari Kinnunen[1], PhD; Tinja Lääveri[2,4], MD, PhD

[1]Department of Health and Social Management, Faculty of Social Sciences and Business Studies, University of Eastern Finland, Kuopio, Finland

[2]Department of Computer Science, School of Science, Aalto University, Espoo, Finland

[3]Department of Social Work, Faculty of Social Sciences, University of Lapland, Rovaniemi, Finland

[4]Infectious Diseases, Inflammation Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

Corresponding Author:
Susanna Martikainen, PhD
Department of Health and Social Management, Faculty of Social Sciences and Business Studies, University of Eastern Finland, Kuopio, Finland

Abstract

Background: The integration of health care and social welfare services together with the consolidation of health care information systems (HISs) and client information systems (CISs) has become a timely topic. Despite this development, there is a scarcity of systematic research on physicians', registered nurses' (RNs) and social welfare professionals' (SWPs) experiences of participating in the development of HISs and CISs.

Objective: This study aimed to examine how physicians, RNs and SWPs experience collaboration with HIS or CIS vendors, and what kinds of end users have participated in HIS or CIS development.

Methods: National cross-sectional usability surveys were conducted in Finland among RNs and SWPs in 2020 and physicians in 2021. Questions concerning participation experiences were analyzed by professional group, working sector, managerial position, and age.

Results: In total, 4683 physicians, 3610 RNs, and 990 SWPs responded to the surveys. In all 3 professional groups, those working in nonmanagerial positions and the youngest respondents participated least in HIS or CIS development, and 76% (n=3528) of physicians, 78% (n=2814) of RNs and 67% (n=664) of SWPs had not participated at all. When comparing the groups, physicians were least aware of feedback processes and least satisfied with vendors' interest in end-user feedback and the manner and speed of HIS development. Those who had dedicated working time for HIS or CIS development were less critical of vendors' interest and responsiveness to development ideas than those who had not participated at all. In all 3 professional groups, the youngest were most dissatisfied with HIS and CIS vendor collaboration.

Conclusions: Experiences of participation in HIS and CIS development were relatively negative across all 3 professional groups, with physicians being the most critical. Dialogue and collaboration between developers and end users—also the youngest ones and frontline workers—need improvement; simply increasing allotted working time is unlikely to produce more positive participation experiences.

(JMIR Hum Factors 2025;12:e51495) doi:10.2196/51495

KEYWORDS

participation; development; usability; user experience; physician; nurse; social worker; information system; national survey; system development; users; user feedback; cross-sectional survey; Finland; Finnish

Introduction

Background

The increasing collaboration between health care and social care and the consequent need for integrated information systems (ISs) warrants studies on the end-user participation viewpoints of all major user groups. Physicians [1-5], nurses [1,2], and social welfare professionals (SWPs) [6] are not satisfied with the usability or daily work support of health care information systems (HISs) and client information systems (CISs) [7]. Furthermore, these professionals experience collaboration with HIS and CIS vendors and developers to be unsatisfactory. However, the majority would be willing to participate in

information system development, but suitable means are lacking [1,2,6,8-11].

End users' participation in IS development is considered to lead to a better user experience and increased user acceptance [12,13]. However, the characteristic inherent complexity of HISs and CISs complicates the implementation of many user participation methods [14,15]. Although end-user participation is regarded as essential in HIS and CIS development [16], without careful management throughout the software development process, participation alone does not guarantee system success [17-19]. End-user participation may even cause more problems than benefits, particularly if the ISs are expected to solve organizational problems [20]. Users may experience that their participation does not affect IS functionalities in a desired manner [1,4,6,16]. In addition, emphasizing administrative information needs and wishes instead of the needs of frontline professionals can complicate work processes by increasing the requirements for data entry and thus impair the workflows [21]. IS users in leadership positions have different needs for ISs than frontline professionals [22].

Comprehensive HISs and CISs are complex ISs used by dozens of user roles in a wide variety of use contexts [23,24]. Many countries, including Finland, are integrating health and social welfare services and consequently ISs [25-27]. One of the rationales behind this development is that those with high numbers of visits to health care often need social welfare services and vice versa [28]. From the point of view of patients and clients, treatment and service packages often include both health care and social services [29-33], which emphasizes the need for fluent information exchange between professional groups to guarantee high-quality and safe care. Consequently, end users from the major professional groups are needed in the IS development processes.

Context of the Study: Health Care, Social Welfare and HISs and CISs in Finland

Until 2023, municipalities (n=309 in 2022) were responsible for organizing social welfare services and primary health care (health centers) in Finland. A total of 20 hospital districts, jointly owned by the municipalities of the region, organized specialized medical care; 5 university hospitals provided tertiary care. Although one-third of outpatient visits to physicians are to private providers (eg occupational health care), the variety of services provided by the private sector is narrow; for example, there are no private intensive care units (ICUs) or labor and delivery units [34]. In social welfare, municipalities or federations of municipalities often purchase some services from private service providers and non-governmental organizations (n=3971 in 2017) [27]. In 2018, there were 19,627 working-age physicians, 70,198 RNs, and 34,523 SWPs in Finland [35-37].

In Finland, the first HISs and CISs were implemented in the 1970s [27]. While only every tenth US hospital used electronic health records (EHRs) as late as in 2010 [38], in Finnish public health care, HIS coverage had already reached 100% by 2007, and by 2014 CISs covered almost all public social services [39,40]. By contrast, in 2020, a quarter of nonpublic social welfare organizations still operated on paper [27,41].

All public hospitals and health centers had joined the Kanta services (national patient data repository and electronic prescription system) by 2015 [42,43], but implementation of the national data repository for social welfare services only began in 2020 [44]. This has required considerable resources from both health care and social welfare organizations and IS vendors over the years [45].

During 2020‐21, in public health care, 2 leading specialized care and 2 leading primary care EHR brands were in wide use. In addition, 4 EHR brands covered both primary and specialized care, of which 1 also covered tertiary care (including functionalities for eg, operating rooms, ICUs, radiology, and emergency departments) and 6 out of the 7 nationally defined social welfare service lines. In addition, 1 EHR brand covered most private sector health care. In public social welfare, 2 CIS brands were in use in most municipalities [46]. In addition, EHR brands were also used in social welfare [41]. In 2018‐21, a new IS was deployed in Southern Finland with 47,000 end users.

Research Questions

In this study, we examined the experiences of physicians, registered nurses (RNs) and SWPs of participating in HIS or CIS development. The data were gathered in 3 large Finnish national surveys in 2020 and 2021. The research questions were as follows:

1. What experiences do physicians, RNs and SWPs have of collaboration with HIS and CIS vendors?
2. Do participation experiences vary by managerial position, employment sector, or age?
3. What types of physicians, RNs and SWPs have participated in HIS and CIS development?

Methods

Survey

This study was part of large national cross-sectional HIS and CIS usability surveys conducted among SWPs and RNs in 2020, and physicians in 2021 [47]. The survey questionnaires and data are available online [47]. The surveys were based on the validated National Usability-Focused HIS Scale [48] and included a section on end users' experiences of participation in HIS or CIS development (Table 1). The statements were originally created and piloted for the national physician survey in 2010 [4,5], and the same statements have been used in later surveys for physicians [1], for RNs [8,49], and SWPs [6]. The survey method and the questionnaire have been described in detail previously [4,5,48].

Table . Questionnaire statements.

Questions and statement/option designations		Statements and options
Question (1) What has been your experience of providing feedback on the information systems you use and their development? please assess the following statements based on your experience. Response options: Fully agree / Somewhat agree / Neither agree nor disagree / Somewhat disagree / Fully disagree		
	Statement A	I know how and to whom I can send feedback about the system if I wish to do so.
	Statement B	The system vendor is interested in feedback about the system provided by the end users.
	Statement C	The system vendor implements corrections and change requests according to the suggestions of the end users.
	Statement D	Corrections and change requests are implemented within a reasonable time frame.
Question (2) Have you participated in information systems development work?		
	Option A	Yes, some of my working time has been allocated for such development work
	Option B	Yes, in addition to my work
	Option C	No

The link to the survey questionnaire was sent to the members of the Finnish Medical Association (FMA) (>90%) (email address available, n=19,142), RN members of the Finnish Nursing Association, the National Association of Health and Welfare Professionals, and the National Professional Association (n=58,276), and SWPs with at least a Bachelor's degree who were members of the following trade unions: the Union of Professional Social Workers, the Trade Union for the Public and Welfare Sectors, or the Social Science Professionals union (n=12,471) [9,50,51] .

The section addressing end-user participation experiences in HIS or CIS development (Table 1) was identical for the physicians and the SWPs with 5-point Likert scale response statements (fully agree; somewhat agree; neither agree nor disagree; somewhat disagree; or fully disagree). However, for the 2020 RNs' survey, a sixth option "Prefer not to respond / Don't know" was added. Furthermore, unlike in the physicians' and SWPs' surveys, it was not possible for RNs to not respond at all. Up to 25% of RNs chose this sixth option. To make the surveys more comparable, we formed a sixth category also for physicians and SWPs of those who had not responded to the statements. In addition, for the descriptive statistics in Figures 1-4 and Table S1 in Multimedia Appendix 1, we combined the responses "Fully agree" and "Somewhat agree" to form a new category "Agree" and included those from the sixth category in the denominator.

The statement on having allocated time to participate in IS development was identical in all 3 surveys, and the background questions included in this study were all optional in all 3 surveys, so nonrespondents were not included in these numbers.

Figure 1. Participation experiences of Finnish physicians, RNs and SWPs by leadership position (leaders versus others). RN: registered nurse; SWP: social welfare professional.

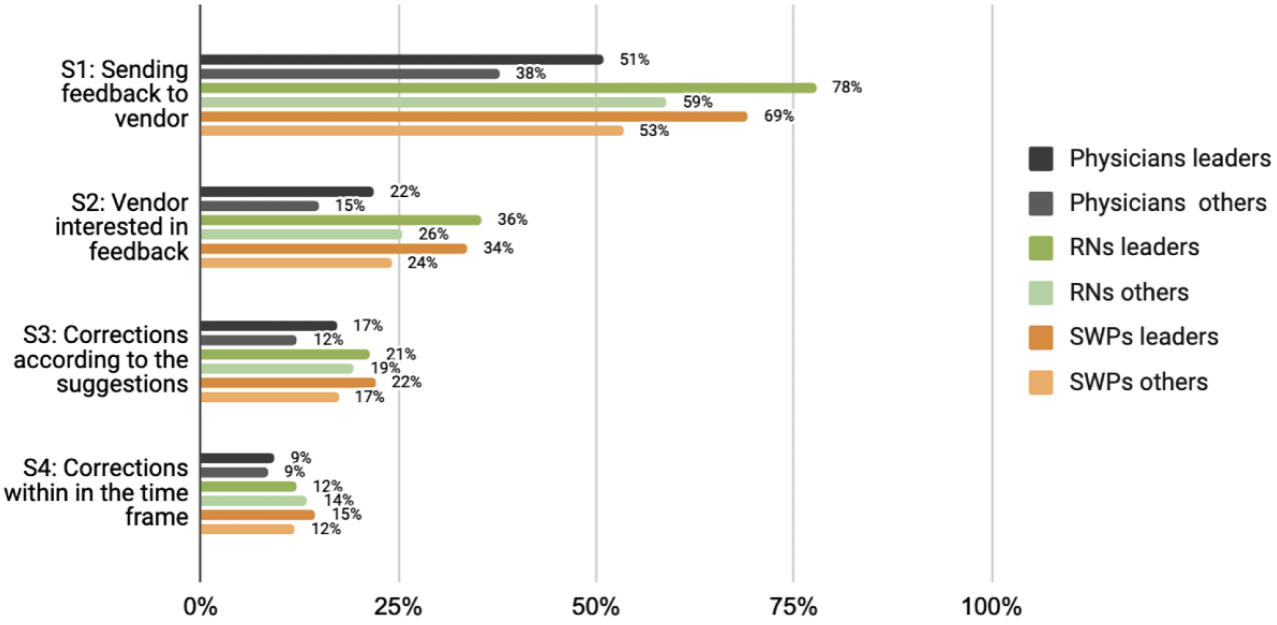


Figure 2. Participation experiences of Finnish physicians, RNs and SWPs by working sector. RN: registered nurse; SWP: social welfare professional.

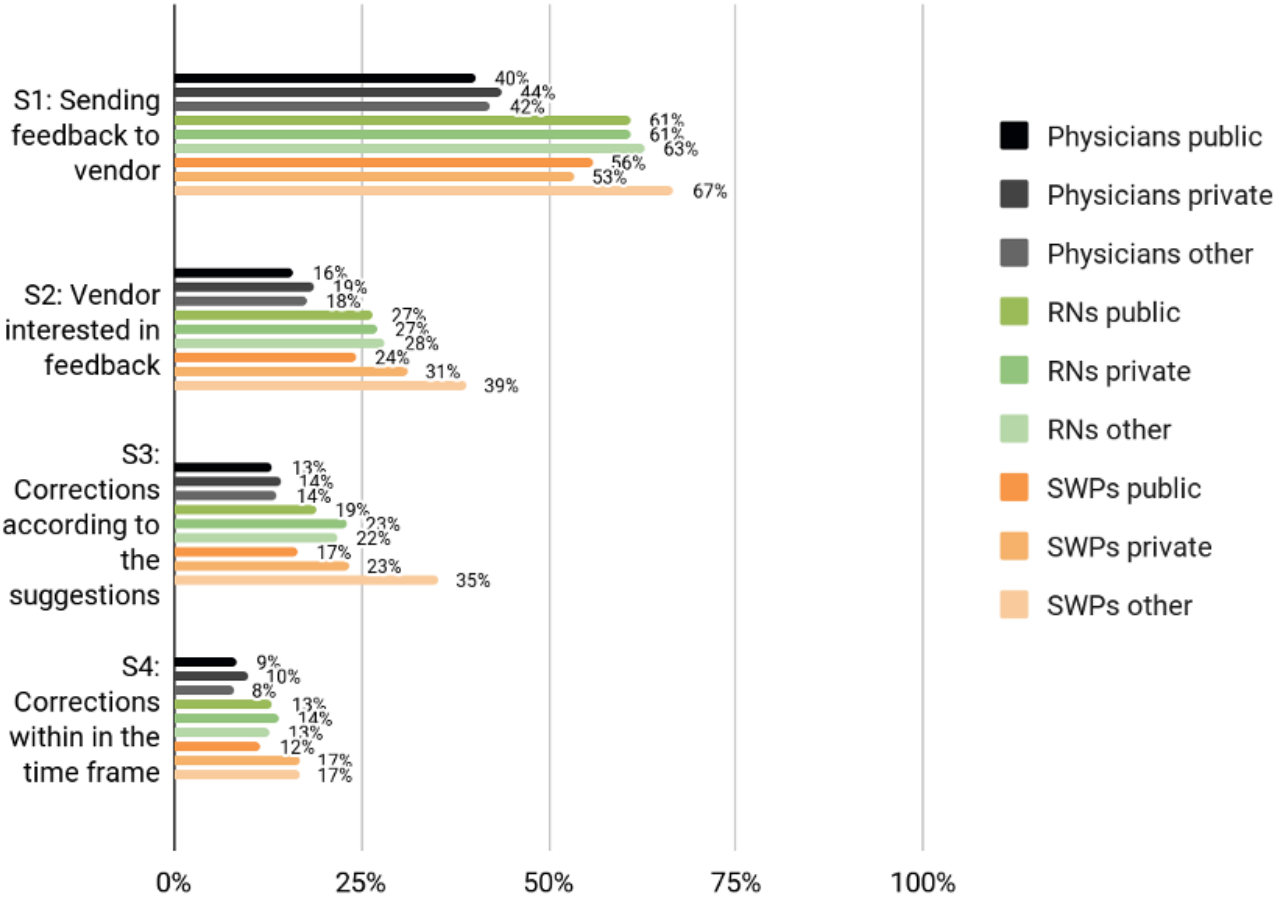


Figure 3. Participation experiences of Finnish physicians, RNs and SWPs by age group. RN: registered nurse; SWP: social welfare professional.

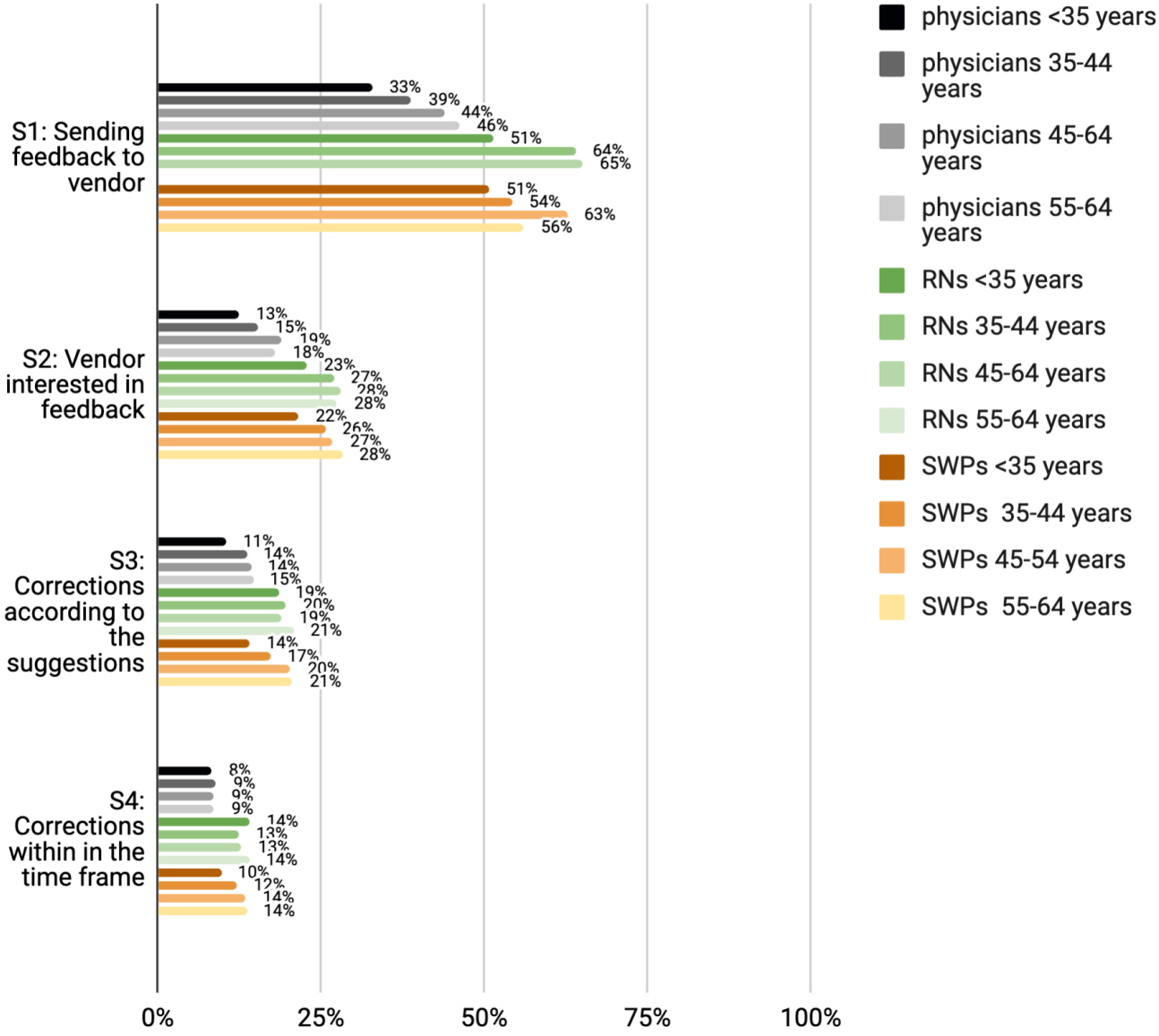
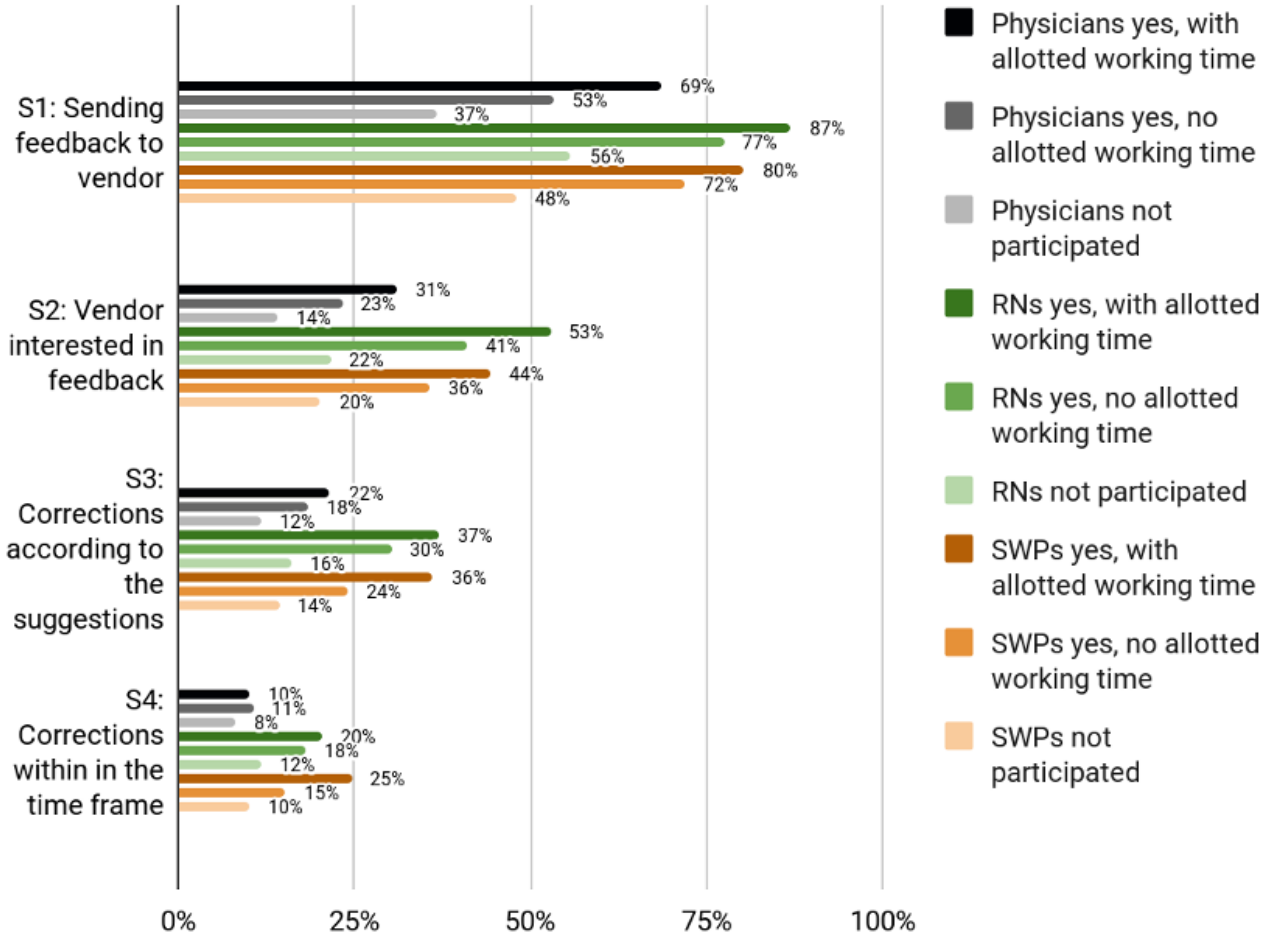


Figure 4. Participation experiences of Finnish physicians, RNs and SWPs by working time allotted for participation. RN: registered nurse; SWP: social welfare professional.



Statistical Analyses

Statistical analyses were carried out with SPSS 28 (IBM Corp). The χ^2 test or Fisher exact test was used as appropriate. Statistical significance was determined as $P<.05$.

Ethical Considerations

According to the national ethical instructions for research, the studies did not require ethical approval (Finnish Advisory Board on Research Integrity 2023) [52].

The autonomy of research subjects was respected, there was informed consent, no harm was possible to the participants and confidentiality of the subjects, and research data were protected. The researchers were not able to identify individual respondents. However, as the data for the RN and SWP studies were collected by a national authority (Finnish Insititute for Health and Welfare), the ethical approval (THL482/6.02.01/2020) was provided by its institutional review board.

Results

Respondent Characteristics

The demographics of the respondents to all 3 surveys are provided in Table 2. In 2021, 4683/19,142 physicians (24.5% of email invitation recipients) participated in the survey, and in 2020, 3610/58,276 RNs (6.2% of email invitation recipients) and 990/12,471 SWPs (7.9% of the theoretical target group) participated in the survey [35-37].

Table . Respondent characteristics.[a]

	Physicians (n=4683), n (%)	Registered nurses (n=3610), n (%)	Social welfare professionals (n=990), n (%)
Working sector			
Public sector	3654 (78)	3076 (85.2)	846 (85.5)
Private	775 (16.5)	456 (12.6)	90 (9.1)
Other	253 (5.4)	78 (2.2)	54 (5.5)
Age group (years)			
<35	949 (20.3)	739 (20.5)	185 (18.7)
35 - 44	1215 (25.9)	833 (23.1)	346 (34.9)
45 - 54	1161 (24.8)	1108 (30.7)	260 (26.3)
55 - 64	1315 (28.1)	921 (25.5)	198 (20)
Leadership position			
Works in a leading or managerial position	1139 (24.3)	406 (11.2)	172 (17.4)
Works in other positions	3543 (75.7)	3204 (88.8)	818 (82.6)

[a]The email invitation was received by 19,142 physicians, 58,276 registered nurses, and 12,471 social welfare professionals. Of these, 4683 (24.5%) physicians, 3,610 (6.2%) registered nurses, and 990 (7.9%) social welfare professionals responded to the survey.

Participation Experiences of the 3 Professional Groups

Physicians appeared to be least knowledgeable of how and where to send feedback, with only 41% (1920/4683) agreeing with statement 1, as compared to 61% (2204/3610) of RNs and 56% (556/990) of SWPs (Figure 5). In terms of vendors' interest in end-user feedback and the manner and speed of IS development, physicians were also least satisfied, with 16% (776/4683), 14% (672/4683), and 9% (407/4683) of them, respectively, agreeing with statements 2, 3, and 4, as compared to 26% (960/3610), 19% (707/3610), and 13% (481/3610), respectively, of RNs and 25% (225/990), 19% (180/990), and 13% (122/990), respectively, of SWPs (Figure 5). In total, 76% (3528/4683) of physicians, 78% (2814/3610) of RNs and 67% (664/990) of SWPs had not participated at all in HIS or CIS development (Figure 6).

Figure 5. Experiences of physicians, RNs and SWPs of collaboration with health care information system and client information system developers. RN: registered nurse; SWP: social welfare professional.

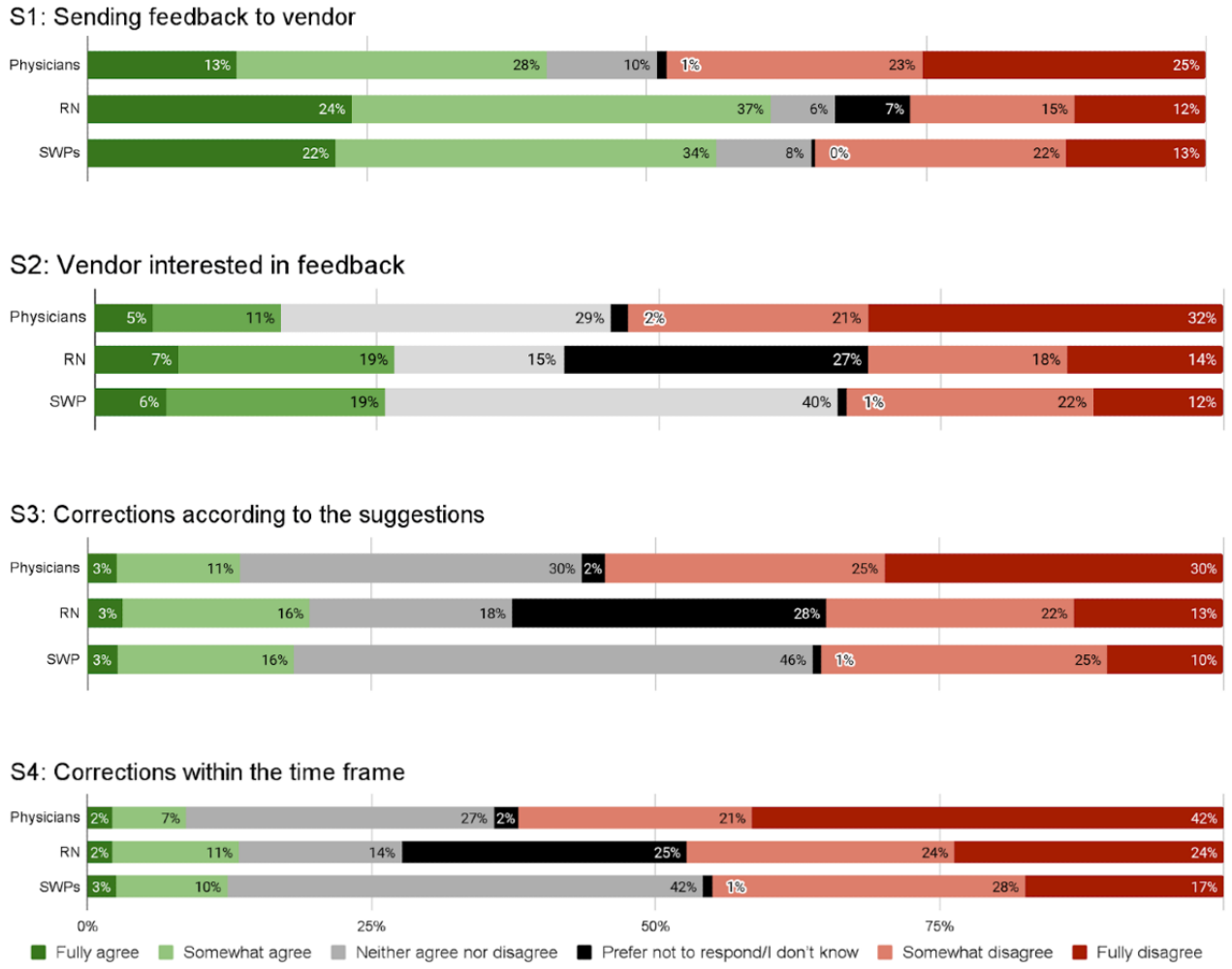
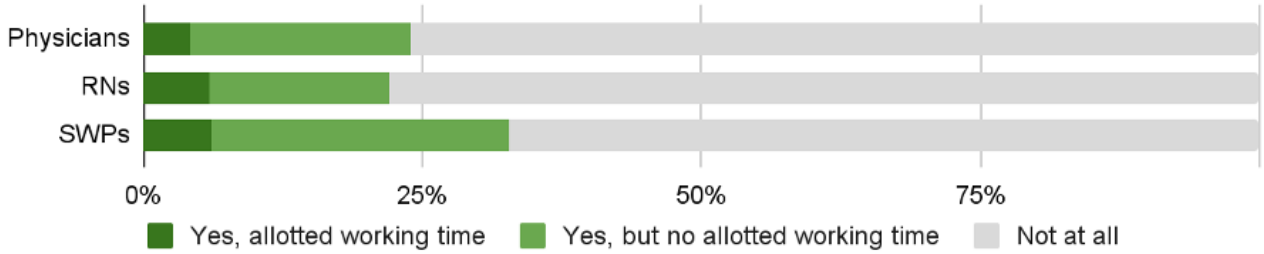


Figure 6. Participation in information system development work by professional group. RN: registered nurse; SWP: social welfare professional.



Factors Associated With Participation Experiences

Leaders, particularly among RNs, were more aware of how and where to send system feedback (statement 1) than those working in nonmanagerial positions (Figures 1-4). Leaders in all 3 professional groups were more satisfied with the system vendor collaboration (statements 2 - 4) than the others. The working sector did not impact end users' experiences. In all professional groups, the youngest appeared least aware of how and where to send feedback and least satisfied with the collaboration. Those who had participated in HIS or CIS development considered IS vendors more interested in feedback and were more satisfied with the manner and speed of system improvements and corrections.

Factors Associated With Having Participated in HIS or CIS Development

In all 3 professional groups, leaders had participated more in IS development than their colleagues in nonleadership positions (Figure 6).

Among physicians and RNs, but not SWPs, those working in the private sector had participated less than their public sector colleagues (Table 3). In all 3 professional groups, the youngest had participated the least (Table 3).

Table . Participation in development by allocated working time.

Participated in development	Yes, allotted working time, n (%)[a]	Yes, but no allotted working time, n (%)[a]	Not at all, n (%)[a]
Leadership position			
Physicians			
Leaders	85 (7.5)	403 (35.7)	640 (56.7)
Others	115 (3.3)	514 (14.6)	2887 (82.1)
RNs[b]			
Leaders	54 (13.3)	131 (32.3)	221 (54.4)
Others	162 (5.1)	449 (14)	2593 (80.9)
SWPs[c]			
Leaders	11 (6.4)	83 (48.3)	78 (45.3)
Others	50 (6.1)	180 (22.1)	586 (71.8)
Working sector			
Physicians			
Public	161 (4.4)	764 (21)	2705 (74.5)
Private	26 (3.4)	104 (13.5)	639 (83.1)
Other	13 (5.3)	49 (20)	183 (74.7)
RNs			
Public	188 (6.1)	500 (16.3)	2388 (77.6)
Private	23 (5)	65 (14.3)	368 (80.7)
Other	5 (6.4)	15 (19.2)	58 (74.4)
SWPs			
Public	53 (6.3)	218 (25.8)	573 (67.9)
Private	6 (6.7)	24 (26.7)	60 (66.7)
Other	2 (3.7)	21 (38.9)	31 (57.4)
Age group (years)			
Physicians			
Age group<35	20 (2.1)	93 (9.9)	827 (88)
Age group 35 - 44	59 (4.9)	227 (18.8)	920 (76.3)
Age group 45 - 54	68 (5.9)	299 (25.9)	787 (68.2)
Age group 55 - 64	52 (4)	289 (22.2)	962 (73.8)
RNs			
Age group<35	34 (4.6)	89 (12)	616 (83.4)
Age group 35 - 44	53 (6.4)	134 (16.1)	646 (77.6)
Age group 45 - 54	83 (7.5)	186 (16.8)	839 (75.7)
Age group 55 - 64	45 (4.9)	166 (18)	710 (77.1)
SWPs			
Age group<35	13 (7.1)	32 (17.5)	138 (75.4)
Age group 35 - 44	23 (6.6)	97 (28)	226 (65.3)
Age group 45 - 54	13 (5)	84 (32.3)	163 (62.7)
Age group 55 - 64	12 (6.1)	50 (25.3)	136 (68.7)

[a]Denominators for calculating percentages are the sum of n values for each row

[b]RN: registered nurse

[c]SWP: social welfare professional

Discussion

Overview

To our knowledge, this is the first study to assess how the major professional groups in the health care and social welfare sector, that is, physicians, RNs, and SWPs, view HIS and CIS development participation. The responses were analyzed by managerial position, employment sector, and age group. Furthermore, we examined which types of professionals have participated in HIS and CIS development.

RNs and SWPs Highly Aware of to Whom and How to Send Development-Related Feedback

The majority of RN (2204/3610, 61%) and SWP (556/990, 56%) respondents knew how and to whom to send development-related feedback; the respective proportion for physicians was 41% (1920/4683). The difference may be explained by mentoring or superuser and training programs during and after HIS implementations among RNs [9,53]. On the other hand, since the response rate among RNs and SWPs was relatively low, those who responded were probably more interested in ISs and thus more aware of HIS and CIS development than those who did not. The findings concur with our earlier studies, which have shown that physicians tend to be more critical towards their HIS compared with RNs and SWPs [2,5].

Leadership and its competence play an important role in the implementation of ISs and investment of resources in digitalization [22,54].

In this study, leaders in all professions and those with allotted working time for HIS or CIS development were more aware of feedback processes than others. These are usually responsible for the orientation of personnel, furthermore, they are often the ones to whom other personnel report development ideas or problems with HIS or CIS use. In all 3 professional groups the youngest were least aware of the feedback processes; whereas the youngest often have mentors who also help with HIS or CIS related problems, it is also probable that the currently available means are not suitable for the younger generations.

Leaders Critical About Cooperation

Although leaders had more often dedicated working time and participated more often in the development of HISs or CISs than the others, they were critical about cooperation with vendors. Their information needs, informatics competencies, and partly also the ISs differ from those working in nonmanagerial positions [55]. The particularly negative viewpoints of physician leaders may be impacted by most of them regularly using HISs for direct patient care, unlike RN or SWP managers who mainly use HISs or CISs for managerial purposes [56]. The most recent study shows that RN managers are able to use HISs for their managerial duties. Due to poor system integration, they need to gather data from different systems for management, which wastes resources inefficiently [22].

Those With Dedicated Working Time Less Dissatisfied With Vendor Cooperation

Those who had dedicated working time for HIS or CIS development were less dissatisfied with vendors' interest and responsiveness to development ideas than those who had not participated at all. They are likely to be more aware of the development processes and timelines of their respective HISs or CISs. Furthermore, since they have been chosen by their respective organizations as participants in HIS or CIS development, their ideas are more likely to become realized. Earlier studies have also found that user participation increases acceptance and active use of HISs contributes to the acceptance and increased active use of HISs [57]. Although not all development ideas are suitable for execution and not all end users can be expected to spend considerable working time on HIS or CIS development, to achieve better engagement in the use of ISs, users need to experience that they are heard and understood [57].

The Youngest Least Satisfied With Vendor Cooperation

Similar to the findings of our previous studies [1,4,6], the youngest were the most dissatisfied with vendors' interest in feedback. This is a particularly important finding as it suggests that the current ways of engaging professionals will not become more suitable or even acceptable to future generations. Barchielli et al [58] also found that younger nurses rely on their colleagues' opinions of health technology use, while older nurses rely on their own experiences.

What Kinds of Users Have Participated in HIS or CIS Development?

Previous studies have shown that impactful participation in IS development requires dedicated working time [58]. Of those working in nonmanagerial positions, 72% - 82% responded that they have not participated at all in HIS or CIS development, whereas the respective proportion for leaders was 45% - 57%. Physician and RN leaders were most likely to have allotted working time for HIS development. Our findings agree with several studies suggesting that managerial viewpoints are likely to become overrepresented in IS development [59,60]. Although the data produced by the ISs is essential for leadership and management purposes, if the participating leaders are not engaged in clinical work or practice the solutions may end up not supporting the needs of frontline workers [61,62].

Health care professionals working in the private sector participated less than their public sector colleagues, among SWPs the differences were minimal. As the majority of Finnish private sector physicians work as private practitioners, their participation would usually result in decreased earnings. It is also likely that the lack of most complex patients in private healthcare reduces the need for HIS development.

In all 3 professional groups, the youngest participated the least. This may be because they are at the stage of learning the clinical content of their work and their employers may not want to invest their time in HIS or CIS development. Khairat et al [62] also

report underrepresentation of physicians in specialization training in HIS development groups. Although not possessing advanced professional skills, the youngest are not burdened with old, often paper-based workflows, which could assist in redesigning processes and enhance the use of newer technologies [62].

How to Improve Satisfaction in HIS and CIS Development

The development processes of large-scale complex systems such as HIS and CIS are typically dominated by cooperative activities involving multiple stakeholders [63-65]. Different information needs must be identified, prioritized, and communicated clearly enough to the IS designers and developers [65].

It remains challenging to increase user input and select appropriate participants and human-centered design methods through the different phases of the participatory development cycle [15,20]. Identifying other factors that influence user experiences, such as decisions made by regulators, policymakers, and administrators, may assist in developing better HISs and CISs [66]. Previous studies have recognized the importance of clinical informaticists who also use HISs or CISs in clinical work in communicating end users' needs and feedback to designers and developers [6,67-75]. From the organizational perspective, the benefits can be seen beyond the IS implementation phase [76]: informatics competent social and health care professionals have been found to be able to improve patient safety and patient care outcomes [67]. In social welfare, however, this role is still being developed [74,75].

Limitations

Our study has some limitations. First, compared with physicians, the lower response rates among RNs and SWPs are likely to have resulted in the selection of more involved and interested participants among these professional groups.

Second, the response rates were highest for physicians. The FMA, which was responsible for collecting the physician data, has a long history of conducting surveys among physicians, the results of which are used in FMA policies. However, the number of participating professionals from all 3 professional groups in the national level studies can be considered high compared to other similar studies [50].

The RNs questionnaire respondents, comprising nurses from various sectors including hospitals, health centers, private practice, and social care, were found to align with the target population according to Statistics Finland's employment data for nurses, midwives, and community nurses [9].

During data collection with SWPs questionnaire, incomplete contact information in membership registries limited survey outreach. To compensate, survey invitations were distributed also through sector networks and social media. The study's final sample comprised 990 SWP respondents, with an estimated response rate of 8%. It's important to note this limitation when interpreting findings. Nonetheless, the sample size was considerable and exhibited diversity in age, service backgrounds, and geographical representation across Finland [51].

Third, our questionnaire did not cover how end users were involved in HIS or CIS development and whether the means of participation would impact satisfaction with the process and the end results. Further research is needed on best practices of user participation in the development of complex HIS and CIS systems.

Conclusion

The fluent use of HISs and CISs is a prerequisite for efficient and safe health care and social welfare, as currently professionals spend a considerable part of their working time with ISs and rely on them as their primary source of information. User participation of all major professional groups—physicians, RNs and SWPs—and their involvement in development are essential for the success of complex HIS or CIS. Compared with RNs and SWPs, physicians appeared to be more critical towards IS vendors and the success of participatory HIS development. As even those with allotted working time were mostly dissatisfied with vendor cooperation, it is evident that simply allocating more end users' working time for HIS and CIS development will not guarantee satisfaction; rather, dialogue between end users and developers needs improvement. New means are needed to better engage all end-user groups, particularly the youngest ones and those working in nonmanagerial positions.

Acknowledgments

This study was part of larger projects supported by the Strategic Research Council of the Academy of Finland (projects 303607 and 327145) and the Ministry of Social Affairs and Health (project 112241). We thank the Finnish Medical Association, the University of Oulu, University of Eastern Finland, the Union of Health and Social Care Professionals in Finland, the Finnish Nurses Association, the National Association of Health and Welfare Professionals, the National Professional Association, the Union of Professional Social Workers, the Trade Union for the Public and Welfare Sectors, and the Social Science Professionals Union for their collaboration. We would also like to thank the respondents to the questionnaires. According to the local and national ethical instructions for research (Finnish Advisory Board on Research Integrity) instructions, this study did not require ethical approval.

Authors' Contributions

SM was responsible for the main supervision, writing and visualization of the research output. TL was mainly responsibile for the statistical analysis. All authors contributed to the analysis and writing of the research output. In addition, SM, JV, and TL

contributed to the design of methodology of all 3 studies, SS to the design, data curation and investigation of the SWP study, and UMK to the design of the registered nurse study.

Conflicts of Interest

SM has been previously employed by 2 of the health care information system (HIS) and client information system (CIS) software vendors included in the study. SS has been previously employed by a publicly owned in-house HIS and CIS vendor included in the study. TL has been previously employed by a publicly owned in-house HIS and CIS software vendor included in the study. The employers did not provide any support, financial or otherwise, to the study. Furthermore, these vendors were not involved in the design of the study or in the collection, analysis or interpretation of the data.

Multimedia Appendix 1
Experiences of participation in system development.
[DOCX File, 20 KB - humanfactors_v12i1e51495_app1.docx]

References

1. Martikainen S, Kaipio J, Lääveri T. End-user participation in health information systems (HIS) development: physicians' and nurses' experiences. Int J Med Inform 2020 May;137:104117. [doi: 10.1016/j.ijmedinf.2020.104117] [Medline: 32179254]
2. Kaipio J, Kuusisto A, Hyppönen H, Heponiemi T, Lääveri T. Physicians' and nurses' experiences on EHR usability: comparison between the professional groups by employment sector and system brand. Int J Med Inform 2020 Feb;134:104018. [doi: 10.1016/j.ijmedinf.2019.104018] [Medline: 31835158]
3. Kaipio J, Lääveri T, Hyppönen H, et al. Usability problems do not heal by themselves: national survey on physicians' experiences with EHRs in Finland. Int J Med Inform 2017 Jan;97:266-281. [doi: 10.1016/j.ijmedinf.2016.10.010] [Medline: 27919385]
4. Martikainen S, Viitanen J, Korpela M, Lääveri T. Physicians' experiences of participation in healthcare IT development in Finland: willing but not able. Int J Med Inform 2012 Feb;81(2):98-113. [doi: 10.1016/j.ijmedinf.2011.08.014] [Medline: 21956004]
5. Viitanen J, Hyppönen H, Lääveri T, Vänskä J, Reponen J, Winblad I. National questionnaire study on clinical ICT systems proofs: physicians suffer from poor usability. Int J Med Inform 2011 Oct;80(10):708-725. [doi: 10.1016/j.ijmedinf.2011.06.010] [Medline: 21784701]
6. Martikainen S, Salovaara S, Ylönen K, et al. Social welfare professionals willing to participate in client information system development - results from a large cross-sectional survey. Inform Health Soc Care 2022 Oct 2;47(4):389-402. [doi: 10.1080/17538157.2021.2010736] [Medline: 34877899]
7. Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of electronic health record use with physician fatigue and efficiency. JAMA Netw Open 2020 Jun 1;3(6):e207385. [doi: 10.1001/jamanetworkopen.2020.7385] [Medline: 32515799]
8. Kinnunen UM, Hyppönen H, Liljamo P, Saranto K. Nurses' experiences of health and social care information systems. In: Vehko T, Ruotsalainen S, Hyppönen H, editors. E-Health and E-Welfare of Finland: National Institute for Health and Welfare (THL); 2019:130-147.
9. Kinnunen UM, Heponiemi T, Rajalahti E, Ahonen O, Korhonen T, Hyppönen H. Factors related to health informatics competencies for nurses-results of a national electronic health record survey. Comput Inform Nurs 2019 Aug;37(8):420-429. [doi: 10.1097/CIN.0000000000000511] [Medline: 30741730]
10. Topaz M, Ronquillo C, Peltonen LM, et al. Nurse informaticians report low satisfaction and multi-level concerns with electronic health records: results from an international survey. AMIA Annu Symp Proc 2016;2016:2016-2025. [Medline: 28269961]
11. Melnick ER, West CP, Nath B, et al. The association between perceived electronic health record usability and professional burnout among US nurses. J Am Med Inform Assoc 2021 Jul 30;28(8):1632-1641. [doi: 10.1093/jamia/ocab059] [Medline: 33871018]
12. Barki H, Hartwick J. Measuring user participation, user involvement, and user attitude. MIS Q 1994 Mar;18(1):59. [doi: 10.2307/249610]
13. Kujala S. User involvement: a review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
14. Lin WT, Shao BBM. The relationship between user participation and system success: a simultaneous contingency approach. Inf & Manag 2000 Sep;37(6):283-295. [doi: 10.1016/S0378-7206(99)00055-5]
15. Kushniruk A, Nøhr C. Participatory design, user involvement and health IT evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
16. Bano M, Zowghi D. A systematic review on the relationship between user involvement and system success. Inf Softw Technol 2015 Feb;58:148-169. [doi: 10.1016/j.infsof.2014.06.011]

17. Bano M, Zowghi D, da Rimini F. User satisfaction and system success: an empirical exploration of user involvement in software development. Empir Softw Eng 2017 Oct;22(5):2339-2372. [doi: 10.1007/s10664-016-9465-1]
18. Johnson L. Effective organizational communication for large-scale healthcare information technology initiatives. In: McCormick KA, Gugerty B, Mattison JE, editors. Healthcare Information Technology Exam Guide for CHTS and CAHIMS Certification: McGraw Hill Education; 2018.
19. Rahimi B, Karimian S, Ghaznavi A, Heydarlou MJ. Development and evaluation of dental image exchange and management system: a user-centered perspective. arXiv. Preprint posted online on Jun 4, 2022. [doi: arXiv:2206.01966]
20. Gillingham P. Electronic information systems in human service organisations: the what, who, why and how of information. Br J Soc Work 2015 Jul;45(5):1598-1613. [doi: 10.1093/bjsw/bcu030] [Medline: 27559213]
21. Gillingham P. Electronic information systems and human services organisations: avoiding the pitfalls of participatory design. Br J Soc Work 2015 Mar 1;45(2):651-666. [doi: 10.1093/bjsw/bct126] [Medline: 27559213]
22. Saranto K, Koponen S, Vehko T, Kivekäs E. Nurse managers' opinions of information system support for performance management: a correlational study. Methods Inf Med 2023 Jun;62(S 01):e63-e72. [doi: 10.1055/a-1978-9727] [Medline: 36379471]
23. Sittig DF, Wright A, Coiera E, et al. Current challenges in health information technology-related patient safety. Health Informatics J 2020 Mar;26(1):181-189. [doi: 10.1177/1460458218814893] [Medline: 30537881]
24. Gillingham P. Developments in electronic information systems in social welfare agencies: from simple to complex. Br J Soc Work 2019 Jan 1;49(1):135-146. [doi: 10.1093/bjsw/bcy014]
25. Vilpponen H, Grundström M, Abrahamsson P, University of Jyväskylä, Finland and City of Kerava, Finland. Combining social service and healthcare as the first country in the world: exploring the impacts on information systems. J Adv Infor Tech 2018;9(4):84-88. [doi: 10.12720/jait.9.4.84-88]
26. Atherton IM, Lynch E, Williams AJ, Witham MD. Barriers and solutions to linking and using health and social care data in Scotland. Br J Soc Work 2015 Jul;45(5):1614-1622. [doi: 10.1093/bjsw/bcv047]
27. Vehko T, Ruotsalainen S, Hyppönen H, editors. E-health and e-welfare of Finland. Check point 2018. Julkari.: National Institute for Health and Welfare (THL); 2019. URL: https://www.julkari.fi/bitstream/handle/10024/138244/RAP2019_7_e-health_and_e-welfare_web_4.pdf?sequence=4&isAllowed=y [accessed 2025-01-06]
28. Hunt KA, Weber EJ, Showstack JA, Colby DC, Callaham ML. Characteristics of frequent users of emergency departments. Ann Emerg Med 2006 Jul;48(1):1-8. [doi: 10.1016/j.annemergmed.2005.12.030] [Medline: 16781914]
29. Smits FTM, Mohrs JJ, Beem EE, Bindels PJE, van Weert HCPM. Defining frequent attendance in general practice. BMC Fam Pract 2008 Apr 15;9(21):21. [doi: 10.1186/1471-2296-9-21] [Medline: 18412954]
30. Vedsted P, Christensen MB. Frequent attenders in general practice care: a literature review with special reference to methodological considerations. Public Health 2005 Feb;119(2):118-137. [doi: 10.1016/j.puhe.2004.03.007] [Medline: 15694959]
31. Social determinants of health. World Health Organization. 2022. URL: https://www.who.int/health-topics/social-determinants-of-health#tab=tab_1 [accessed 2025-01-06]
32. Jørgensen JT, Andersen JS, Tjønneland A, Andersen ZJ. Determinants of frequent attendance in Danish general practice: a cohort-based cross-sectional study. BMC Fam Pract 2016 Jan 28;17(9):9. [doi: 10.1186/s12875-016-0412-4] [Medline: 26821807]
33. Report to Congress; social risk factors and performance under medicare's value-based purchasing programs. Office of the Assistant Secretary for Planning and Evaluation. 2016. URL: https://aspe.hhs.gov/reports/report-congress-social-risk-factors-performance-under-medicares-value-based-purchasing-programs [accessed 2025-01-06]
34. Social welfare and health care system in Finland, responsibilities. Ministry of Social Affairs and Health. 2019. URL: https://stm.fi/en/social-and-health-services/responsible-agencies [accessed 2025-01-06]
35. Lääkärityövoima [Web page in Finnish]. Finnish Medical Association. 2022. URL: https://www.laakariliitto.fi/laakariliitto/tutkimus/laakarityovoima/ [accessed 2025-01-06]
36. Terveys- ja sosiaalipalvelujen henkilöstö 2021 [Web page in Finnish]. Finnish Institute for Health and Welfare. URL: https://thl.fi/fi/tilastot-ja-data/tilastot-aiheittain/sosiaali-ja-terveydenhuollon-resurssit/terveys-ja-sosiaalipalvelujen-henkilosto [accessed 2025-01-06]
37. Terveys- ja sosiaalipalvelujen henkilöstö 2018: tuleva sote-uudistus koskee 7,3 prosenttia työssäkäyvistä [Web page in Finnish]. Julkari. 2021. URL: https://urn.fi/URN:NBN:fi-fe2021121460373 [accessed 2025-01-06]
38. Non-federal acute care hospital electronic health record adoption. Office of the National Coordinator for Health Information Technology. 2017. URL: https://www.healthit.gov/data/quickstats/non-federal-acute-care-hospital-electronic-health-record-adoption [accessed 2025-01-06]
39. Kuusisto-Niemi S, Hyppönen H, Ruotsalainen S. Availability and use of e-welfare in finland. In: Vehko T, Ruotsalainen S, Hyppönen H, editors. E-Health and E-Welfare of Finland Check Point 2018: National Institute for Health and Welfare (THL); 2019. URL: https://www.julkari.fi/bitstream/handle/10024/138244/RAP2019_7_e-health_and_e-welfare_web_4.pdf?sequence=4&isAllowed=y [accessed 2025-01-06]
40. Kärki J, Ryhänen M. Use of Information and Communications Technology in Social Services in 2014: National Institute for Health and Welfare (THL); 2015.

41. Salovaara S, Silén M, Vehko T, Kyytsönen M, Hautala S. Use of Information and Communication Technology in Social Welfare Services in 2020: National Institute for Health and Welfare (THL); 2021.
42. Citizens. Kanta. 2022. URL: https://www.kanta.fi/en/citizens [accessed 2025-01-06]
43. Jormanainen V. Large-scale implementation and adoption of the Finnish national Kanta services in 2010–2017: a prospective, longitudinal, indicator-based study. FinJeHeW 2018;10(4):381-395. [doi: 10.23996/fjhw.74511]
44. Client data repository for social welfare services. Kanta. 2022. URL: https://www.kanta.fi/en/professionals/client-data-archive-for-social-welfare-services [accessed 2025-01-06]
45. Jormanainen V, Reponen J. CAF and CAMM analyses on the first 10 years of national Kanta services in Finland. FinJeHeW 2020;12(4):302-315. [doi: 10.23996/fjhw.98548]
46. Jormanainen V, Rötsä M, Parhiala K. Social care client information systems in Finnish municipality social services in 2017. Finn J EHealth E Welfare 2019;11(1-2):125-138. [doi: 10.23996/fjhw.76835]
47. Implementation of the monitoring of information system services in social welfare and healthcare. Finnish Institute for Health and Welfare. 2023. URL: https://thl.fi/en/topics/information-management-in-social-welfare-and-health-care/information-management-in-finland/follow-up-of-the-information-system-services-in-social-welfare-and-health-care/implementation [accessed 2025-01-06]
48. Hyppönen H, Kaipio J, Heponiemi T, et al. Developing the National Usability-Focused Health Information System Scale for physicians: validation study. J Med Internet Res 2019 May 16;21(5):e12875. [doi: 10.2196/12875] [Medline: 31099336]
49. Kaihlanen AM, Gluschkoff K, Saranto K, Kinnunen UM. The associations of information system's support and nurses' documentation competence with the detection of documentation-related errors: Results from a nationwide survey. Health Informatics J 2021;27(4):14604582211054026. [doi: 10.1177/14604582211054026] [Medline: 34814758]
50. Viitanen J, Reponen J, Lääveri T, et al. Potilastietojärjestelmissä on yhä paljon parannettavaa: lääkärien kokemukset käytettävyydestä vuosilta 2010-2021. Usability of electronic health record systems needs improvements – physicians' experiences from 2010-21. Finn Med J 2024(41):79 [FREE Full text]
51. Salovaara S, Hautala S, Silén M. Sosiaalialan ammattilaisten kyvykkyys tietojärjestelmien käyttäjinä ja kokemukset tietojärjestelmien tuesta työlle [Article in Finnish]. Fin Je He W 2022;14(2):208-225. [doi: 10.23996/fjhw.110119]
52. Finnish Advisory Board on Research Integrity (TENK). URL: https://tenk.fi/en/tenk [accessed 2025-01-13]
53. Rhodes B, Short A, Shaben T. Effectiveness of training strategies that support informatics competency development in healthcare professionals. In: Shachak A, Borycki EM, Reis SP, editors. Health Professionals' Education in the Age of Clinical Information Systems, Mobile Computing and Social Networks: Academic Press; 2017:299-322. [doi: 10.1016/B978-0-12-805362-1.00015-2]
54. Booth RG, Strudwick G, McBride S, O'Connor S, Solano López AL. How the nursing profession should adapt for a digital future. BMJ 2021;373:1190. [doi: 10.1136/bmj.n1190]
55. Hübner U, Shaw T, Thye J, et al. Technology Informatics Guiding Education Reform – TIGER: an international recommendation framework of core competencies in health informatics for nurses. Methods Inf Med 2018;57(Suppl 1):e30-e42. [doi: 10.3414/ME17-01-0155] [Medline: 29956297]
56. Backonja U, Langford LH, Mook PJ. How to support the nursing informatics leadership pipeline: recommendations for nurse leaders and professional organizations. Comput Inform Nurs 2022 Jan 1;40(1):8-20. [doi: 10.1097/CIN.0000000000000827] [Medline: 34996883]
57. Vahteristo A, Jylhä V. Effects of user participation in the development of health information systems on their evaluation within occupational health services. In: Integrated Citizen Centered Digital Health and Social Care: IOS Press; 2020:207-211. [doi: 10.3233/SHTI200724]
58. Barchielli C, Marullo C, Bonciani M, Vainieri M. Nurses and the acceptance of innovations in technology-intensive contexts: the need for tailored management strategies. BMC Health Serv Res 2021 Jul 3;21(1):639. [doi: 10.1186/s12913-021-06628-5] [Medline: 34215228]
59. Gillingham P. From bureaucracy to technocracy in a social welfare agency: a cautionary tale. Asia Pac J Soc Work Dev 2019 Apr 3;29(2):108-119. [doi: 10.1080/02185385.2018.1523023]
60. Devlieghere J, Roose R. Documenting practices in human service organisations through information systems: when the quest for visibility ends in darkness. Soc Inclu 2019;7(1):207-217. [doi: 10.17645/si.v7i1.1833]
61. Høstgaard AM, Bertelsen P, Nøhr C. Methods to identify, study and understand end-user participation in HIT development. BMC Med Inform Decis Mak 2011 Sep 28;11(1):1-11. [doi: 10.1186/1472-6947-11-57] [Medline: 21955493]
62. Khairat S, Burke G, Archambault H, Schwartz T, Larson J, Ratwani RM. Focus section on health IT usability: perceived burden of EHRs on physicians at different stages of their career. Appl Clin Inform 2018 Apr;09(02):336-347. [doi: 10.1055/s-0038-1648222]
63. Saleem N, Steel D, Gercek G, Chandra A. Significance of user participation in a hospital information system success. Health Care Manag 2017;36(2):199-205. [doi: 10.1097/HCM.0000000000000156]
64. Sommerville I, Cliff D, Calinescu R, et al. Large-scale complex IT systems. Commun ACM 2012 Jul;55(7):71-77. [doi: 10.1145/2209249.2209268]
65. Dalsgaard P, Frich J, Riisgaard T. Four challenges of designing digital services to support participatory design. Presented at: Nordic Human-Computer Interaction Conference; Oct 8, 2022; Aarhus Denmark. [doi: 10.1145/3546155.3547273]

66. Tutty MA, Carlasare LE, Lloyd S, Sinsky CA. The complex case of EHRs: examining the factors impacting the EHR user experience. J Am Med Inform Assoc 2019 Jul 1;26(7):673-677. [doi: 10.1093/jamia/ocz021] [Medline: 30938754]
67. Lagsten J, Andersson A. Use of information systems in social work – challenges and an agenda for future research. Eur J Soc Work 2018 Nov 2;21(6):850-862. [doi: 10.1080/13691457.2018.1423554]
68. AMIA's role in CIS and history. American Medical Informatics Association. URL: https://amia.org/careers-certifications/clinical-informatics-subspecialty/amias-role-cis-and-history [accessed 2025-01-13]
69. Singer JS, Cheng EM, Baldwin K, Pfeffer MA, UCLA Health Physician Informaticist Committee. The UCLA Health Resident Informaticist Program - a novel clinical informatics training program. J Am Med Inform Assoc 2017 Jul 1;24(4):832-840. [doi: 10.1093/jamia/ocw174] [Medline: 28115427]
70. Sieja A, Markley K, Pell J, et al. Optimization sprints: improving clinician satisfaction and teamwork by rapidly reducing electronic health record burden. Mayo Clin Proc 2019 May;94(5):793-802. [doi: 10.1016/j.mayocp.2018.08.036] [Medline: 30824281]
71. Lehmann CU, Gundlapalli AV, Williamson JJ, et al. Five years of clinical informatics board certification for physicians in the United States of America. Yearb Med Inform 2018 Aug;27(1):237-242. [doi: 10.1055/s-0038-1641198] [Medline: 29681038]
72. Cummins MR, Gundlapalli AV, Murray P, Park HA, Lehmann CU. Nursing informatics certification worldwide: history, pathway, roles, and motivation. Yearb Med Inform 2016 Nov 10;25(1):264-271. [doi: 10.15265/IY-2016-039] [Medline: 27830261]
73. Conrad JB, Magsamen‐Conrad K. Understanding the impact of the coronavirus pandemic on families involved in the child welfare system: technological capital and pandemic practice. Child Fam Soc Work 2022 Feb;27(1):11-21. [doi: 10.1111/cfs.12876]
74. Parker-Oliver D, Demiris G. Social work informatics: a new specialty. Soc Work (Stell) 2006 Apr 1;51(2):127-134. [doi: 10.1093/sw/51.2.127]
75. Kuusisto-Niemi S, Saranto K, Rissanen S. Informatics in social services: research, developments, and outcomes from the finnish perspective. In: Tavana M, Ghapanchi AH, Talaei-Khoei A, editors. Healthcare Informatics and Analytics Emerging Issues and Trends: IGI Global; 2015:1-26. [doi: 10.4018/978-1-4666-6316-9.ch002]
76. HIMSS Nursing Informatics Workforce Survey. HIMSS. 2020. URL: https://www.himss.org/resources/himss-nursing-informatics-workforce-survey [accessed 2025-01-06]

Abbreviations

CIS: client information system
FMA: Finnish Medical Association
HIS: health care information system
IS: information system
RN: registered nurse
SWP: social welfare professional

Edited by A Kushniruk; submitted 02.08.23; peer-reviewed by A Sieja, W Tesfaye; revised version received 02.05.24; accepted 08.11.24; published 22.01.25.

Please cite as:
Martikainen S, Viitanen J, Salovaara S, Kinnunen UM, Lääveri T
Comparisons of Physicians', Nurses', and Social Welfare Professionals' Experiences With Participation in Information System Development: Cross-Sectional Survey Study
JMIR Hum Factors 2025;12:e51495
URL: https://humanfactors.jmir.org/2025/1/e51495
doi:10.2196/51495

Original Paper

Patient-Generated Collections for Organizing Electronic Health Record Data to Elevate Personal Meaning, Improve Actionability, and Support Patient–Health Care Provider Communication: Think-Aloud Evaluation Study

Drashko Nakikj[1], PhD; David Kreda[1], BA; Karan Luthria[1], BCS; Nils Gehlenborg[1], PhD

Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Nils Gehlenborg, PhD
Department of Biomedical Informatics
Harvard Medical School
10 Shattuck Street
Suite 514
Boston, MA, 02115
United States
Phone: 1 8572725075
Email: nils@hms.harvard.edu

Abstract

Background: Through third party applications, patients in the United States have access to their electronic health record (EHR) data from multiple health care providers. However, these applications offer only a predefined organization of these records by type, time stamp, or provider, leaving out meaningful connections between them. This prevents patients from efficiently reviewing, exploring, and making sense of their EHR data based on current or ongoing health issues. The lack of personalized organization and important connections can limit patients' ability to use their data and make informed health decisions.

Objective: To address these challenges, we created Discovery, an experimental app that enables patients to organize their medical records into collections, analogous to placing pictures in photo albums. These collections are based on the evolving understanding of the patients' past and ongoing health issues. The app also allows patients to add text notes to collections and their constituent records. By observing how patients used features to select records and assemble them into collections, our goal was to learn about their preferred mechanisms to complete these tasks and the challenges they would face in the wild. We also intended to become more informed about the various ways in which patients could and would like to use collections.

Methods: We conducted a think-aloud evaluation study with 14 participants on synthetic data. In session 1, we obtained feedback on the mechanics for creating and assembling collections and adding notes. In session 2, we focused on reviewing collections, finding data patterns within them, and retaining insights, as well as exploring use cases. We conducted reflexive thematic analysis on the transcribed feedback.

Results: Collections were useful for personal use (quick access to information, reflection on medical history, tracking health, journaling, and learning from past experiences) and clinical visits (preparation and raising physicians' awareness). Assembling EHR data into reliable collections could be difficult for typical patients due to considerable manual work and lack of medical knowledge. However, automated collection building could alleviate this issue. Furthermore, having EHR data organized in collections may have limited use. However, augmenting them with patient-generated data, which are entered with flexible richness and structure, could add context, elevate meaning, and improve actionability. Finally, collections might produce a misconstrued health picture, but bringing the physician in the loop for verification could increase their clinical validity.

Conclusions: Collections can be a powerful tool for advancing patients' proactivity, awareness, and self-advocacy, potentially facilitating patient-centered care. However, patients need better support for incorporating their own everyday data and adding meaningful annotations for future reference. Improvements in the comprehensiveness, efficiency, and reliability of the collection assembly process through automation are also necessary.

(JMIR Hum Factors 2025;12:e50331) doi:10.2196/50331

KEYWORDS

mobile health; patients; electronic health records; sensemaking; data organization; collections; awareness; proactivity; self-advocacy; patient–health care provider communication

Introduction

Background

Digital technologies play a pivotal role in facilitating patient-centered care that focuses on understanding patients' needs and fostering shared decision-making with their health care providers [1]. However, this approach relies heavily on well-informed patients and on effective patient-provider communication [2]. In the United States, a major step toward meeting these requirements was enabling patients to access their medical records from multiple providers through third-party applications [3,4]. While this is a significant achievement, enabling patients to engage in making sense of these data and turn the insights they gain from this process into actionable steps remains a challenge. Consequently, many patients lack a satisfactory understanding of their health, feel discouraged to self-advocate, and have mediocre communication with their providers, which is at odds with the core values of patient-centered care. Therefore, addressing the challenges around sensemaking and the usability of health data will be important to advancing patient-centered care and empowering patients to take an active role in their health journey.

Making sense of data, or *sensemaking*, is a cyclic process that involves cognitive activities for answering complex questions [5]. These activities involve repeated access to artifacts, identifying relevant information, finding information relationships, and presenting the answers in an understandable format [6]. Patients face a plethora of sensemaking challenges to manage their health. They need to assemble health information from different providers and identify outliers, correlations, and trends to become educated on health topics, drive decision-making, and formulate discussion points with health care providers [7]. Unfortunately, robust platforms to support patients in making sense of their clinical data are lacking.

Several commercial mobile apps such as Apple Health Records [8], iBlueButton [9], OneRecord [10], and 1upHealth [11], along with the academic web application Discovery [12], advance patient sensemaking by offering data visualizations and specialized views for data exploration. These views help patients uncover interesting patterns related to prevalence, periodicity, co-occurrence, and pre-post analysis of medical events. Typically, these solutions organize electronic health record (EHR) data by type, time stamp, or provider. However, such approaches provide very little support for patients in finding deeper connections between their medical records that are required for understanding health issues, reflecting on medical history, or preparing for clinical encounters.

Moreover, these apps do not allow patients to annotate their medical records or save their sensemaking progress, forcing them to remember findings or record them elsewhere. This necessitates patients to revisit the same data repeatedly to refresh inferences and recreate mental notes. Such work is typically tedious and frustrating, leading to anxiety and missing crucial information. Consequently, patients can form skewed health impressions, resulting in poor decisions or risky actions. In clinical visits, the inability to communicate the sensemaking insights to physicians may hinder optimal treatment, leading to repeated tests or medical errors.

To address these limitations, we explored an alternative solution that organizes EHR data into collections based on health issues and ongoing problems [13]. Inspired by findings that a problem-based view of EHR data improves clinician awareness, prioritization, and decision-making in the intensive care unit [14], we adapted a similar approach for patients anchored in the *data-frame* sensemaking theory [15]. Data frames (ie, structured mental models) or *collections* of health data, as we refer to them in our previous work [13], systematically break down problems and help in answering complex questions. These capabilities of the collections hold significant potential for managing health data effectively for all sorts of patients, particularly those with complex medical histories or multiple comorbidities [13]. By organizing abundant data around health issues, collections help patients avoid fragmented health impressions, a common challenge for those with multiple comorbidities. Patients who see multiple specialists can use collections to track the development of specific issues and share insights across providers, raising awareness and improving care coordination.

More precisely, our proposed concept of *collections* allows patients to dynamically organize, adapt, and explore their health information based on evolving needs and available data. For example, a patient managing cardiovascular issues might create a Blood Pressure collection to consolidate related records, which could later branch into more specific collections such as Extreme Blood Pressures or Blood Pressure Lab Work. These refined groupings help uncover patterns and dependencies among factors such as BMI, diet, cholesterol levels, and blood pressure, enhancing understanding and facilitating proactive management of health conditions. Gathering insights from the collections may also help patients have more productive discussions with their providers.

In our study, we extended the capacity to transform EHR data into collections and facilitated reasoning regarding them. As patients' sensemaking of health data is driven by finding outliers, correlations, and trends [7,16], we enabled capabilities to identify data patterns within the collections. In addition, we supported the assembly of relevant medical records for the collections by helping patients visually explore, find temporal patterns [17-20], and make sense of their EHR data within a single, context-preserving view [12]. Acknowledging patient requests for automation [13], we offered manually assembled and system-assembled collections. We also allowed for personal data input through free-text notes, fulfilling previously identified patient needs [13].

A notable advancement in our work is moving from mock-ups [13] to a fully functional mobile app, Discovery. The app

provides patients with a realistic platform to interact with collections, uncovering deeper insights into preferred mechanisms for creating, refining, and using these collections. Moreover, we intended to motivate patients to see collections as tools for self-advocacy during clinical visits, which was identified as a key use case in our earlier research [13].

Objectives

In this study, we used the Discovery app to explore patients' needs, preferences, and desired interactions for organizing EHR data and delve into potential use cases. More concretely, we asked the following research questions (RQs):

- What are the needs and feature preferences for organizing EHR data from multiple providers? (RQ 1)
  - What are the patients' experiences with creating and building collections?
  - How effectively does our app allow patients to organize their EHR data using the concept of collections?
  - How can we better meet patients' needs for meaningful EHR data organization?
- What purpose would organizing the EHR data in collections have? (RQ 2)

To answer these RQs, we conducted a qualitative evaluation study with 14 participants.

Methods

Description of Discovery

General Overview

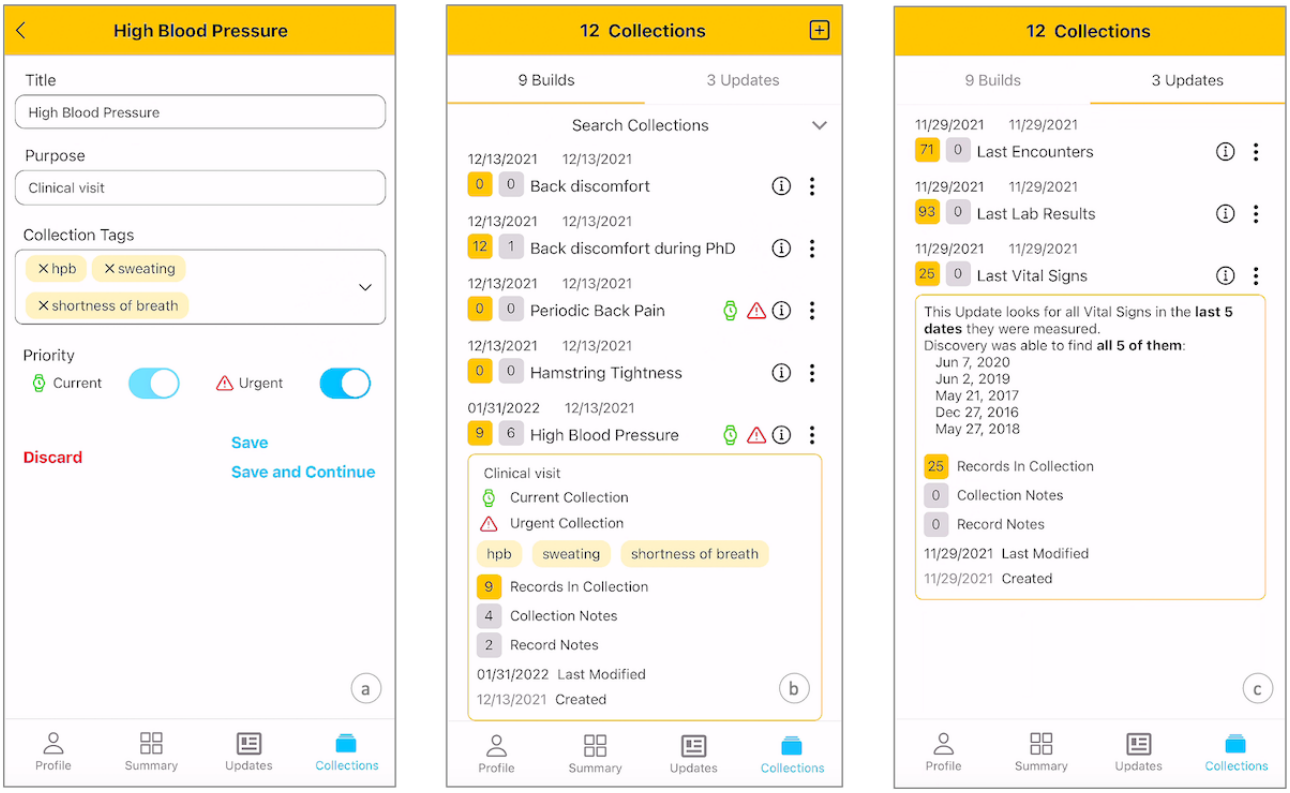
Discovery is a noncommercial iPhone app designed to help patients make sense of their EHR data. It introduces a novel concept that organizes EHR data into personalized, problem-based collections. In addition, it allows patients to add their own observations and insights, providing context and complementing their medical records with patient-generated data (PGD).

For this study, the app was restricted to accessing synthetic EHR data through a Fast Healthcare Interoperability Resources (FHIR) format [21] from the SMART Health IT repository [22]. Discovery accesses only structured EHR data and relies on a 2-level hierarchy. At the highest level, there are the record categories ("Conditions," "Immunizations," and "Vital Signs"). Each record category has multiple record types ("Vital Signs: Body Height," "Body Weight," and "Blood Pressure"), and each record type can have multiple instances ("Blood Pressure: systolic: 125, diastolic: 90," and "date: 02/01/2022"). Each instance of data in our app is called a record and corresponds to an FHIR resource with standardized attributes.

Organizing Records in Collections

The manually created collections are called Builds (creation shown in Figure 1A and list shown in Figure 1B), whereas the app-assembled collections are called Updates (Figure 1C).

Figure 1. (A) Creation of a new collection; (B) list of patient-built collections—Builds; (C) list of system-generated collections—Updates.



Patients can manually create a new collection (Figure 1A) by entering a name and additional metadata. To prioritize and distinguish the collections, we introduced descriptors: purpose, tags, and priority. For example, in Figure 1A, the patient created a High Blood Pressure collection with the purpose Clinical Visit for an upcoming appointment. They also added tags such as

hpb, sweating, and shortness of breath for reference. The priority descriptor indicates that the collection addresses a current and urgent issue. Patients can modify any of this information as the collection develops and changes.

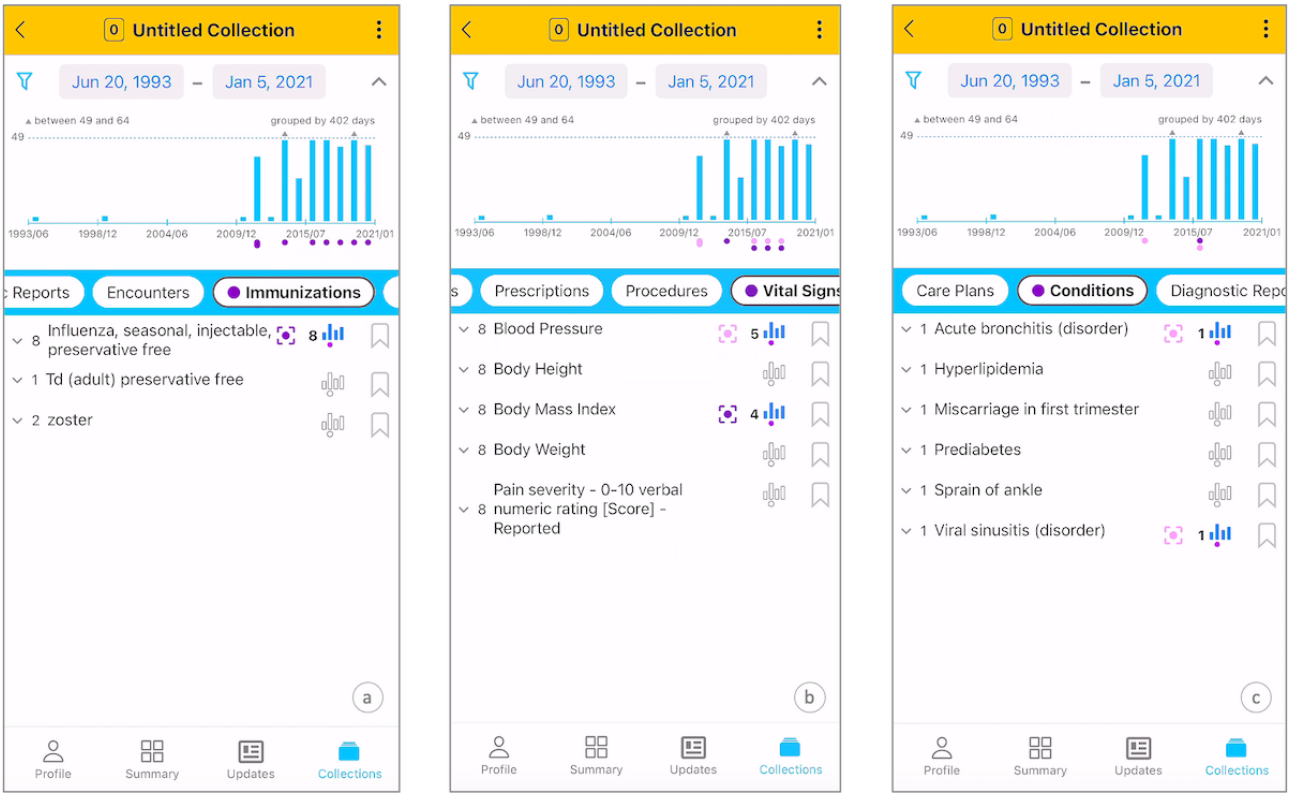
As patients repeatedly create collections, their EHR data transform into a problem-based list, as shown in Figure 1B, which includes health issues such as back discomfort, hamstring tightness, and high blood pressure. Collections are displayed with their name, creation and last modification dates, record count, number of patient-added notes, and labels for current (green clock) and urgent (danger sign) issues. An information button provides a summary with explanations. For example, the High Blood Pressure collection was created on December 31, 2021, and last modified on January 31, 2022. An information panel summarizes its contents: 9 records, 6 notes (4 for the collection and 2 for individual records), and tags for detailed search (hpb, sweating, and shortness of breath).

Updates automatically familiarize patients with the latest relevant events without requiring any action on their part (Figure 1C). The app scans all records and matches predefined templates, such as recent encounters, laboratory test results, and vital signs. In Figure 1C, 3 Updates are listed—Last Encounters, Last Lab Results, and Last Vital Signs—based on the last 5 dates when corresponding records were logged by the provider. The list entry follows the same structure as that of the Builds without the labels for currency and urgency. Although records in an Update cannot be changed, patients can add and remove notes. Patients can also clone an Update into a new Build and rename it if needed, allowing them to reuse and customize the assembled records for specific purposes. The reason for having this distinction between Updates and Builds is to delineate what the system and the user have produced and which entity is responsible for the collection that might have led to certain actions.

Identifying Relevant Records and Patterns for Collections

Discovery offers an interactive visualization to find patterns within records. The Timeline depicts record counts in equal time intervals, showing the prevalence of medical events. A dotted horizontal line marks the threshold above which the volume of records is considered abnormal, corresponding to the mean record count per interval. Gray triangle glyphs indicate values between the mean and 2 SDs, whereas red triangles highlight values of >2 SDs. By highlighting individual records or entire record types, patients can explore patterns that may be saved in existing collections or trigger the creation of new collections. For example, Figure 2A shows periodicity of influenza shots (when and how frequently influenza shots were administered), Figure 2B shows the co-occurrence of high blood pressure and high BMI, and Figure 2C shows the absence of respiratory conditions after an amoxicillin prescription.

Figure 2. Finding medical event patterns: (A) periodicity of influenza shots (showing when and how frequently influenza shots were administered), (B) co-occurrence of high blood pressure and high BMI, and (C) absence of respiratory conditions after an amoxicillin prescription.



The FHIR resources (ie, medical records) are represented with Record Cards, which display the clinical information in human-readable format. Patients can use a Filter and Date Picker to narrow down record categories and time frames for displayed records. Selected record categories appear in a Sliding Tabs control, allowing patients to swipe left or right for immediate access. Clicking on a record category in the Sliding Tabs organizes it by record type, represented with Accordions (eg,

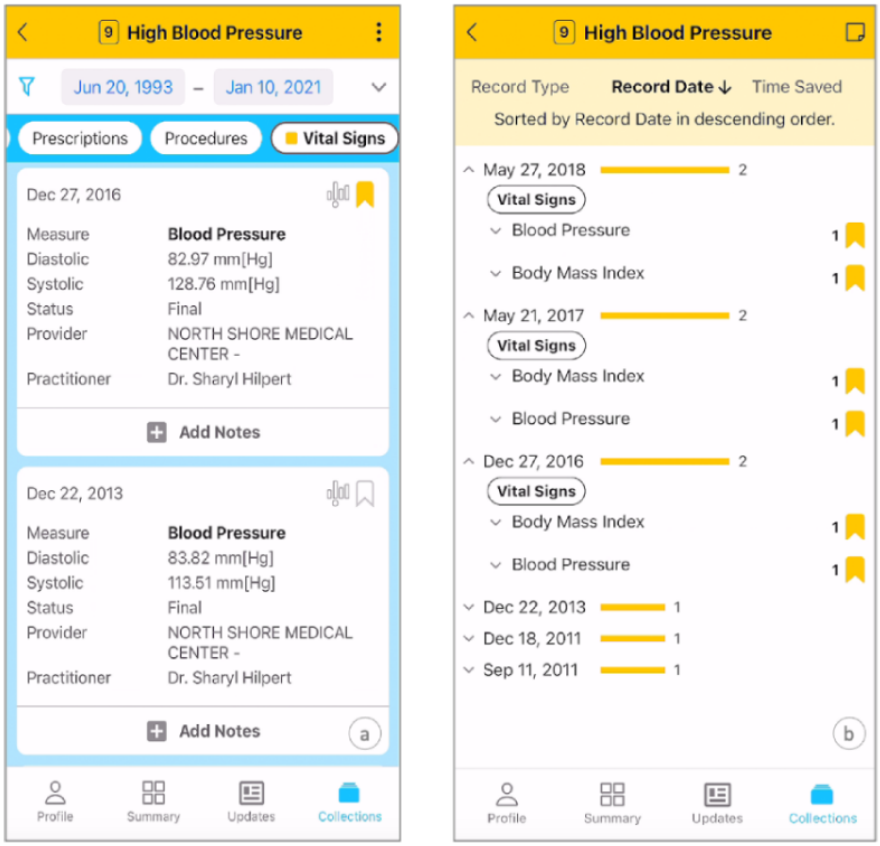
the Accordion for Immunizations category will have 3 sections for the Flu shot, Tdap, and Zooster record types). Patients can expand the Accordion sections to scroll through individual Record Cards. For example, Figure 2A shows 8 Record Cards for the influenza shot under the Immunization Accordion section. Accessing and revisiting records involves swiping the Sliding Tabs and selecting record categories, with Accordions retaining their expanded or collapsed state and scrolling position. This method is more context preserving compared to existing solutions, which require repeated back-and-forth navigation through different views for each record category and record type.

Producing Insights for the Collection

Patients can save individual records or entire record types by tapping the bookmark icon in the corresponding Record Card or Accordion section. Figure 3A demonstrates adding individual Blood Pressure records to a collection using a Record Card. Records can be removed from collections by tapping the selected bookmark again.

To identify data relationships and produce insights, patients can inspect a collection in the Collection Review (Figure 3B). Here, records can be viewed by type, date recorded in the provider's EHR system, or time added to the collection. When sequence matters, records can be ordered chronologically. Patients can also remove records in the Collection Review by tapping the selected bookmark. For example, in Figure 3B, the patient views records grouped by recording dates in descending order, with yellow bars indicating record counts by date. They observe high blood pressure and high BMI co-occurring 3 times, suggesting a pattern that may need further investigation or discussion with a physician.

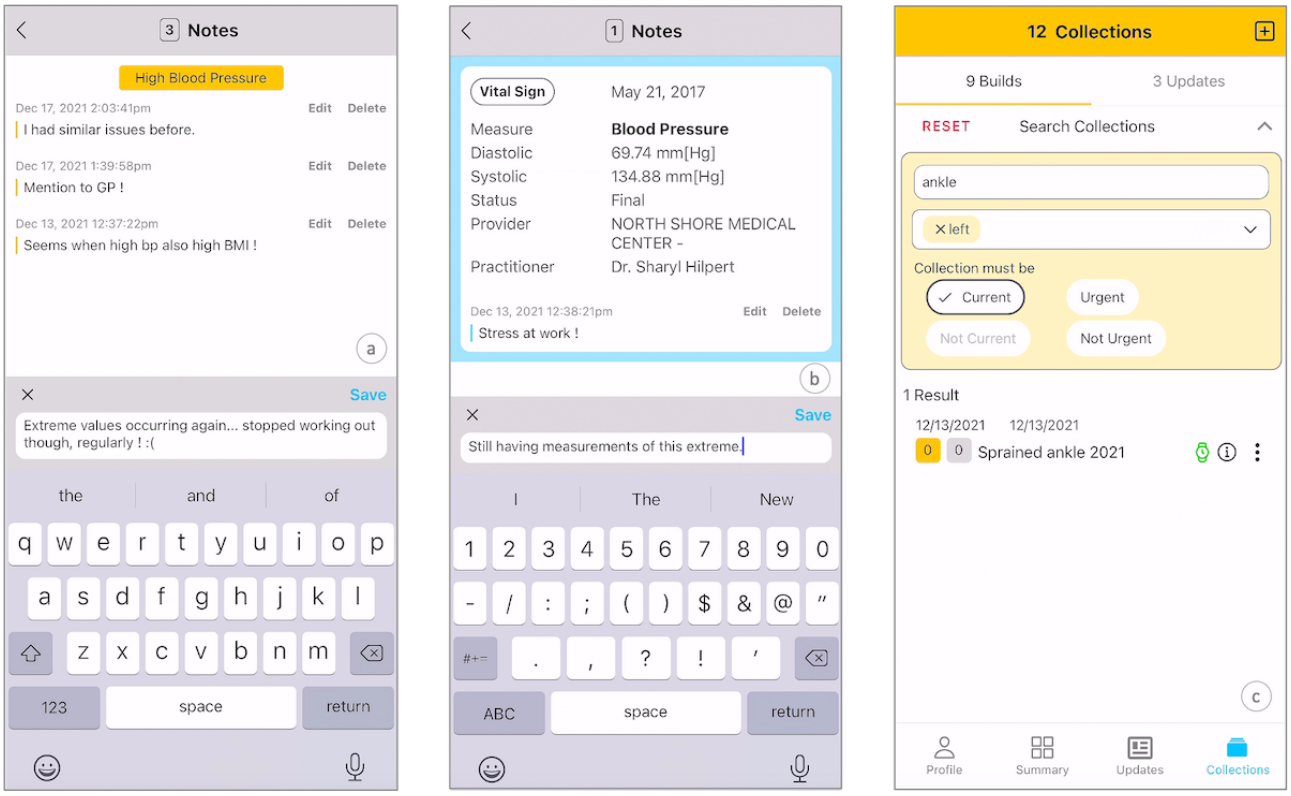
Figure 3. (A) Saving a record to a collection and (B) reviewing the collection and inspecting data patterns.



Supporting PGD

Patients can enter notes for a collection (Figure 4A) or individual records within it (Figure 4B) to add personal insights, progress, observations, details, or disease journal entries. Notes can be modified or removed at any time. In Figure 4A, the first note provides context on past high blood pressure experiences. The second note serves as a reminder to mention occasional high blood pressure to the general practitioner. The currently created note adds context about noticing changes in blood pressure after stopping regular workouts. In Figure 4B, the patient contextualizes a high blood pressure measurement taken during a stressful period at work and notes the recurrence of high values.

Figure 4. (A) Adding a note to a collection, (B) adding a note to a record, and (C) searching the collections.



Searching the Collections

Free-text search targets the collection name, purpose, and notes (Figure 4C), as well as tags and priority. Results dynamically update as the search query is constructed.

Study Design

Participants

We recruited 14 participants from our email list compiled from previous recruiting efforts and Craigslist. This number is sufficient to uncover usability issues and provide rich findings as per current design research practices [23] and literature on user feedback quality [24]. We balanced the sample by age, gender, and medical history (including healthy individuals, those with acute episodes, and those with chronic illnesses). Eligibility criteria included adults fluent in English; possessing an iPhone (iPhone 6 or above) and a laptop or desktop computer (screen size of 13“ or more) with a stable, fast internet connection for both devices; and with normal or corrected vision, no color blindness, and medical records from one or more providers. Medical history was self-reported.

Table 1 illustrates the detailed participant demographics collected using the questionnaire from Table S1 in Multimedia Appendix 1. Our 14 adult study participants included 10 (71%) women and 4 (29%) men aged 24 to 61 years (mean age 35.6, SD 12.6; median 30.5 years). All had some college experience: half (7/14, 50%) held bachelor’s degrees, 14% (2/14) had some graduate experience, and 14% (2/14) had completed master’s degrees. Participants had between 2 and 15 health care providers, with half (7/14, 50%) having ≥6. The 29% (4/14) of the participants who were healthy saw physicians a few times a year. Those with chronic illnesses had been managing their diseases for 1 to 20 years.

All participants were comfortable with daily technology use, and 21% (3/14) had work experience in data analytics. Most (11/14, 79%) used third-party apps to track mental health, weight loss, sleep, and exercise. All but 1 (13/14, 93%) used provider-patient portals to review test and laboratory results, refill prescriptions, and schedule appointments. However, participants found it cumbersome to remember multiple passwords and difficult to find specific information due to interface issues. Data sharing among providers was often slow or impossible, forcing participants to print and assemble records for clinical visits.

Table 1. Participant demographics.

Participant ID	Age (y)	Sex	Educational level	Medical issues	Health care providers, N
P1	61	Male	2-year college	High blood pressure, sleep apnea, fatty liver (stage 3), and liver transplant	7-8
P2	45	Female	Some college	High blood pressure, prediabetes, obesity, and hypothyroidism	6-7
P3	51	Male	Some graduate school (incomplete degree)	Hypertension and gout	2
P4	24	Male	Bachelor's degree	Childhood asthma, recent hernia surgery, and mental health therapy	5
P5	24	Female	Bachelor's degree	No chronic issues	2
P6	27	Female	Bachelor's degree	No chronic issues	3-4
P7	28	Male	Graduate degree	Physical and mental (20 years)	Approximately 13
P8	37	Female	Master's degree	Thyroid condition (many years)	3-5
P9	29	Female	3 years (incomplete degree)	Chronic migraines (10 years) and hemiplegic migraines (1 year)	≥12
P10	59	Female	Bachelor's degree	Yes—5 years	10-15
P11	25	Female	Bachelor's degree	No chronic issues	3-5
P12	33	Female	Master's degree	Anemia (2 years) and IBS[a] (10 years)	3
P13	19	Female	Some college (incomplete degree)	Asthma (since she was a baby) and EDD[b] (10 years)	>5
P14	32	Female	Master's degree	No chronic issues	3

[a]IBS: irritable bowel syndrome.

[b]EDD: eosinophilic digestive disease.

Procedures

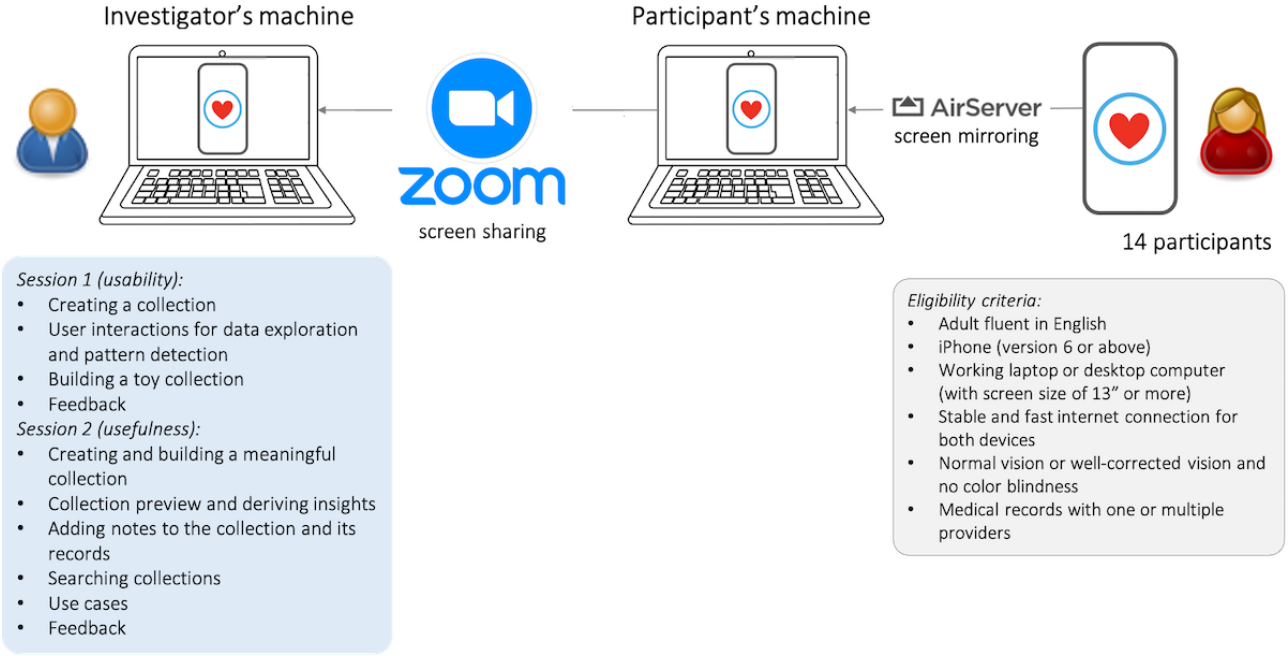
Inspired by existing patient portal usability studies, the study procedures were tailored to our RQs [23,25,26] and are presented in Figure 5. The study was conducted remotely via Zoom (Zoom Video Communications) meetings. Participants downloaded our app on their iPhone using Apple TestFlight for beta testing [27] and mirrored their iPhone screen on their laptop using AirServer (App Dynamic ehf.) [28]. The laptop screen was then shared with the researcher for observation.

Participants attended an initial 60-minute session (session 1) and a follow-up 45-minute session (session 2). Both sessions used the concurrent think-aloud protocol to gather rich qualitative data on user needs, perceptions, and preferences efficiently [29]. Textbox 1 illustrates 2 example topics and detailed tasks for creating a toy collection and realistic collection from sessions 1 and 2, respectively. The full set of 13 topics for both sessions and their detailed tasks that follow a similar structure are shown in Tables S2 and S3 in Multimedia Appendix 1.

In session 1, the researcher first collected demographics and digital health consumer information from participants (Table S1 in Multimedia Appendix 1). The session then evaluated the usability of the app (RQ 1). Participants were introduced to the app's key concepts and features and were then given tasks to learn the interactions for creating and assembling relevant records in collections. After each task block, participants provided feedback. Tasks included creating a toy collection, adding descriptors, exploring scoping mechanisms for record categories and time ranges, navigating records using the Sliding Tabs, and inspecting the interactive Timeline visualization. Participants also completed tasks related to finding periodicity, co-occurrence, and pretest-posttest analysis of medical events using the Timeline; saving records to a collection; and reviewing it. The session concluded with feedback on the intuitiveness, usability, and usefulness of creating, building, and reviewing collections, as well as data exploration and pattern detection features (Textbox 2).

Session 2 focused on the usefulness of collections in real-life settings (RQ 2). Participants performed tasks involving creating a collection tied to a specific issue (eg, high blood pressure) and purpose (eg, clinical visit). After each task block, participants provided feedback. Tasks included creating a High Blood Pressure collection, adding descriptors for a clinical visit, selecting records related to high blood pressure, identifying patterns in selected records, adding notes to the collection and specific records for clinical context, and using the search feature to find other collections. The session concluded with feedback on the usefulness of collections, brainstorming real-life use case scenarios, and suggesting potential improvements (Textbox 2).

Figure 5. Study design—14 participants for 2 one-on-one sessions with a researcher. Session 1 focused on usability, and session 2 examined usefulness. Feedback was collected on how participants would use collections for their own data and needs.



Textbox 1. Example tasks from sessions 1 and 2.

Session 1: creating a collection (5 min)

- For the purposes of learning the basic mechanics around building a Collection from scratch, first Create a new Collection and name it "Toy Collection."
- Now, let's add some descriptors about the Collection.
  - Please add a purpose to the Collection, something related to learning about this app.
  - Now, add a couple of tags to further describe, summarize or annotate the Collection for future quick access.
  - Finally, specify the priority of this Collection by marking it as urgent.
- What was your experience with creating the new Collection?
  - How intuitive was it? Have you seen similar interactions elsewhere?
  - How useful do you find the descriptors for the Collection?
- Now, let's go back to the list of Collections. Let me know how can you see the details about the Collection you just created?
  - How intuitive was it?
- What are possible improvements?

Session 2: building a collection (15 min)

- You will now create a Collection that is more realistic and meaningful for use in a real-life scenario. We will assume that you are preparing for an upcoming visit to your physician's office related to potential issues with high blood pressure. Create a collection called "High Blood Pressure."
  - Add the purpose for the Collection
  - Add a few tags
  - Mark its priority
- Add the Records with blood pressure with systolic value over 120 and BMI over 30.
- What was your experience with assembling the Records for the Collection?
  - How laborious was it?
  - What are some ways in which we can make this assembling process more efficient?
  - How do you feel about having the system prepopulate the Collection for you and let you modify it afterwards?

Textbox 2. Feedback collected at the end of the study sessions.

Session 1 example tasks

- What was your impression of this app?
  - What did you like?
  - What did you dislike?
- How intuitive was the app?
- How easy or hard was it to explore the data?
- How useful were the features in the app to identify patterns in the data?
- How did you like the mechanism for saving Records in the Collection?
- What are some improvements you would like to see?

Session 2 example tasks

- How useful do you think the Collections can be for you?
- What are some use cases for the Collections that you can think of?
- What are some improvements you would like to see for the Collections?
  - Automatic support for building Collections?
  - Automatically finding data patterns in the Collections?
  - Patient-generated data?

Data Collection

We recorded the Zoom meetings for audio and video capture of the entire interaction. The first author also took notes during the meetings.

Data Analysis

Audio recordings were transcribed using Rev (Rev.com, Inc). We analyzed video recordings for additional context and a deeper understanding of participant comments during the think-aloud protocol. Video annotations were added to the transcripts and session notes [23] for reflexive thematic analysis [30]. The first author began by open coding the textual data. Emerging categories were reconciled in meetings with the second and last authors to identify use cases and detailed approaches to organizing and annotating EHR data. These themes were validated in a group meeting with researchers unfamiliar with the collection concept and with our app.

Ethical Considerations

The Harvard Faculty of Medicine Institutional Review Board approved this study (protocol IRB20-1757). Participants signed a consent form, which also allowed them to opt out of the study at any point. Each participant received a US $40 Amazon gift card as compensation. The data obtained from the study sessions did not include any identifiable information about the participants and were stored on a password-protected computer with encryption. Only the research team had access to these data.

Results

Overview

The results from our study are organized and analyzed around five qualitative themes, two applicable to RQ 2 and three applicable to RQ 1: (1) using collections for personal benefit (RQ 2), (2) using collections in a clinical setting (RQ 2), (3) creating and building collections (RQ 1), (4) enhancing collections (RQ 1), and (5) accessing collections (RQ 1). In the remainder of the Results section, we characterize the participants and report on these themes using 16 quotes from 11 different participants. The participants are labeled as P1 to P14.

Purposes for Organizing the EHR Data in Collections

Using Collections for Personal Benefit

Quick Access to Information

Participants perceived the Collections feature as a way to index information at a suitable level of granularity for health issues, topics of interest, or conditions to monitor. They expected that this would give them quick access to current or urgent problems that were being managed:

> *Personally I would like to monitor my asthma because I am using medication for that, the typical inhaler, but I would like to monitor, these days I had certain attacks or shortness of breath, so collections, having it back for that specific condition is very useful to me, because I wouldn't have to speculate about when my last attack was or when my last appointment date was, it's right here for me to access.* [P13]

Reflection on Medical History

Participants viewed collections as a tool to construct and understand their medical history. They expected collections to serve as a repository for the health issues they faced, aiding in reflecting on the existence, prevalence, and development of their health conditions:

> *This will be able to find what happens when I have, in my case, headaches related to my high blood pressure and nothing else, or if you're sick for the flu, influenza and you have fever, it's just related to the influenza, not to the COVID.* [P1]

Keeping Track of Health Status

Most participants envisioned using collections to track their health status proactively. This included monitoring urgent issues needing immediate attention, unstable conditions requiring frequent observation, and treatments needing careful monitoring. Participants also wanted collections to track abnormal laboratory test results and vital sign values across various health issues:

> *Well, for me, it's kind of good [to have medical records organized in collections], especially, for example, blood samples, especially those with high triglycerides or something. Maybe I can collect them and see whether there's a trend for this month, or for January, I'm high in this one. Then second month, I'm also high, so maybe I can lower it down...For collections, just categorize those. Which are high, which are low.* [P5]

Journaling Daily Events

Most of the participants envisioned using collections as a personal diary for coping with diseases and logging measurements and their effects on lifestyle. They wanted to track challenges, successes, and progress toward finding solutions and monitor disease developments:

> *What I've done is I've taken all of my videos and stuff since February. Like I said, I've been to eight different doctors and I've shown them, this is the progress of what's happened from...I've had two surgeries before this surgery where they lanced it, cut it, drained it. Nothing happened. The cyst came back and then it went into my bone. So I'm able to bring these photos, I'm able to bring this timeline, I'm able to bring my frustrations and show this doctor within 30 seconds [using a collection], look, this is what it looked like. And this is my own diary, my own history. It's very important because they're a doctor. They don't know me, they don't know what it looked like on day one.* [P10]

Learning From Past Experiences

Nearly all participants wanted to use collections to identify trends and patterns in their ongoing health experiences, including co-occurring symptoms and treatment effects. They also intended to log food intake, sleep, activity, or stress factors to find triggers for symptoms:

> *I think I could definitely use them there. It's a lot easier now because I could highlight the certain event, and put like my triggers down with it, like migraine on the third was this, you could even put what you took with it. So for me, I would look back and be like, "Oh, I can tell my doctor that, I've had 10 migraines. I took a medication with these three. What's my options." So I think that would be great, it's a great tool that I can actually do that with this app.* [P9]

Using Collections in a Clinical Setting

Preparation for the Clinical Visit

Most participants would use collections to prepare talking points for clinical visits combining personal measurements and notes with medical records from other providers (eg, laboratory and test results). This was crucial because their physicians often did not have access to external data:

> *For collections, I would say that if I'm meeting with multiple providers about one health issue, I could see myself combining all my records there so that multiple providers can see each other's records...I would probably [use the notes], if I needed to jot down a certain time that I took a measurement, or if my doctor told me keep track of what you ate that day, or kind of anything that I would want to have the details for the next time that I go and see the provider.* [P6]

Relying on Collections During Clinical Visits

All participants wanted to share collections with their physicians during clinical visits to establish ground truths, raise awareness of other providers' information, and provide transparent talking points. They believed that adding PGD to collections could describe what happened between visits and raise their physicians' awareness:

> *If you have a condition, you need to check on your blood pressure. You need to communicate that with your doctor, so you can add a note [in the collection] saying like, "Latest, highest blood pressure from this week," from a date.* [P14]

In addition, most participants felt that taking in-visit notes and saving them in a collection could help them understand care plans and take appropriate actions afterward:

> *Maybe [taking notes in the collection for the clinical visit] just for your own personal reference or if you wanted to bring it up later on in another appointment or something, or just maybe, I guess, just general recording of something that happened during that visit.* [P11]

All participants saw physicians as essential partners in reviewing collections and deciding on actions based on their contents. This was primarily because most participants doubted their expertise in determining what should go into collections or which collections to create. While they were very open to include their physicians in the collection curation, some feared that they might overburden physicians with verification inquiries:

> *I guess I would definitely do that [look for patterns in the data and store them in a collection] just*

> *because I can actually consult the doctor. Is this actually correlated or will I have to change my diet because it affects this? At least you can ask the doctor, or confirm whether that is true or not.* [P5]

Needs and Feature Preferences for Organizing EHR Data in Collections

Creating and Building Collections

Manual Workflow for Creating and Building Collections

Overall, all participants expressed satisfaction with the clarity and simplicity of the mechanics to create collections and save records in them. However, those with less medical knowledge and disease experience (relatively healthy and recently diagnosed individuals) thought that initiating and building collections was challenging. For them, it was not always clear what issues deserved separate collections, what records to include in a collection due to delicate dependencies, and why they should invest substantial time and effort in assembling collections. In contrast, the more experienced (chronic) patients were relatively confident in their ability to carve a personalized view of their EHR data. However, some did acknowledge that they might not be as exhaustive and reliable in their data organization.

Participants quickly mastered using the Sliding Tabs with Accordions and appreciated the context-preserving record exploration. However, most of the participants found it challenging to assess the relevance of individual records due to unexplained clinical language and attribute values. Participants requested explanations, prominent visual cues for abnormal values to attract their attention, and time-series visualizations for additional context and noticing trends. They also wanted to be able to sort by date or attribute value, with filtering capabilities that also included adding the filtered records in bulk to a collection.

Some participants desired collections with richer internal structures, recognizing the need to identify subsets of records and their importance within a collection. They suggested linking groups or individual records using specific annotations for easy identification during visual inspection or search:

> *...maybe if certain records are related to each other. So I would want to mark that. And then maybe just have a way of sorting down based on certain labels.* [P8]

Automatic Support for Building Collections

Participants were highly receptive to discussing ideas that could automate building collections. They were interested in seeing their records automatically put into collections based on provenance (provider, physician, hospital, location, and date) and clinical meaning (condition, disease, organ, organ system, and abnormality of the values across records). Some also suggested grouping records into collections based on personal annotations. For all these groupings, they wished to be able to edit the collection manually:

> *If I had neurological problems, neurology collection. If I had urological problems, urology collection. I think that for me at least would seem a more straightforward way to categorize them. But from the categories I've already seen, I find those useful.* [P4]

Several participants suggested using a "seed" to automate collection building, such as naming a collection, adding keywords, or including a few records:

> *And I think a lot of patients don't know where to start, what data to begin with. So if it's something that's already preset, they say, "Okay, I'm suffering from depression or I have diabetes." And the system pulls the different data points that they would need to look at for someone who's diabetic or someone who's dealing with depression, I think that's helpful. Because sometimes, the problem is you don't know where to start and you don't know what to look for.* [P11]

Finally, most of the participants wanted to receive automated help to add or remove records for a collection that had already been created. Few described wanting to choose from a list of suggested records based on the existing content of a collection, revealing records and record patterns otherwise invisible to them. Others saw this automated record offering more as an idea generation approach—needing some follow-up validation, including taking it up with their physicians:

> *When I work, I want to listen to some music and then I'm like, okay. I just don't know what's next. I need something similar to this, the same vibe, but I just can't think about that. And then there is suggestions. And yes, some of it's weird. But maybe like the doctor can also have some help here, and when you review the collections together, they might say, "Hey, listen, this is what the system gave you and that's great. Let's remove a few things. I would suggest you add a couple others. And whatever you put there, it's also fine. And let's keep it that way."* [P6]

Enhancing Collections Through Personally Provided Data

Making Collections Complete

Participants strongly expressed the need to complement their EHR data with daily entries from sensors; self-monitoring devices; and manual measurements of symptoms, treatments, and outcomes in various formats (text, photos, videos, and scanned documents):

> *Actually, I think you can sort of restructure the whole core of the collections on top of two main pillars. The first one would be all of the doctor's data, which is basically hard data, which allows you to diagnose, allows you to run statistical analysis...That could be part of the core data, but all of the context, maybe I'm getting this shortness of breath in my home, watching my TV, might be added by the notes. You have these two types of data. By adding the user data, would allow me to get context, give context, which is important and will allow me to, on a daily basis, keep a record, which in case of data like shortness of breath, I'm having, I'm not having. Would allow the*

> *doctor to have a really unbiased input on symptoms I'm having.* [P7]

Participants wanted to log detailed observations and measurements, pairing treatments with outcomes and symptoms with triggers. They suggested dedicating a special PGD record category for these data, with some preferring complex structures and others favoring simple data entry options:

> *I would probably use the notes quite often just to maybe outline the symptoms I was experiencing and the steps I took to alleviate those symptoms or which doctors I contacted.* [P13]

Making Collections Distinguishable

Participants liked the existing collection descriptors and suggested additional ones. They wanted labels for clarity (clear, unclear, or potential issue), stability (stable or unstable), progress toward resolution, development stages (nonthreatening or threatening), and a list of involved providers and physicians. When collections were related to clinical visits, participants wanted to specify the targeted physicians.

Making Collections Actionable

Participants believed that organizing medical records by health issues in collections was a good start but thought that actionability could be improved with specific insight notes and annotations applied to entire collections:

> *And then as far as the purpose of adding a note to the whole category, I would say that, like you said, if you happen to notice any patterns when you're looking at the data, or basically I would use it for any general or bigger-picture takeaway that I wanted to tell my doctor, "Hey, I noticed this" or something and I wanted to bring it to their attention.* [P6]

Participants envisioned using collection-wide notes to summarize contents or purpose, track progress, describe issue development, and highlight special events. They also wanted notes representing care plans and actions prioritized in a to-do list. Participants intended to use collections to prepare for clinical visits with questions, reminders, and critical measurements. They also saw value in adding collection notes about visit outcomes, key takeaways, and next steps.

Some participants wanted to annotate and highlight keywords or add tags to free-text notes for organized review and pattern identification.

Accessing the Collections

All participants emphasized the importance of fast, reliable access to collections and their contents. They primarily relied on collection descriptors but also desired a deep search feature that would scan through individual records, notes, and annotations within collections.

Discussion

Principal Findings

We identified 3 principal findings of our study. First, participants embraced the collection concept. Unrestrictedly organizing EHR data into collections that map medical records to health issues and track ongoing concerns gave participants a sense of ownership. They felt empowered by developing personalized health narratives that could aid in self-management and communication with their physicians, enhancing their self-advocacy. Second, while participants easily mastered the interface for initiating and adding records to collections, they found the process laborious. They lacked confidence in selecting appropriate records due to limited medical knowledge and requested additional visual cues, explanations, and automatic collection features. There was concern about potential self-misguidance without physician verification. Third, collections would need richer PGD capabilities for adding contextual information not found in participants' EHR data, logging observations, and labeling data. This would enhance the comprehensiveness and accuracy of their health narratives and support foraging, sensemaking, and action taking.

Interpretation of the Findings and Contributions

Overview

On a broader scale, this work contributes to patient-centered care. This is achieved by demonstrating potential to enhance patients' grasp of their health, encourage self-advocacy, and improve patient-provider communication. More accurately, there are several concrete contributions of our work that can be considered as proxies toward achieving the aforementioned objectives: (1) encouraging patient ownership of their EHR data by organizing them into personalized, health issue–based collections; (2) understanding patients' perceptions and preferences for creating, building, and using these collections; and (3) offering design insights for automating collections, integrating rich PGD, enhancing access to collection contents, and using collections to facilitate patient-provider communication.

Going forward, we will situate our findings within a sensemaking framework and discuss contributions related to 3 key patient needs: increasing awareness through independent health sensemaking, proactivity through efficient action taking, and self-advocacy through incorporating evidence-supported patient perspectives into patient-provider communication. We will elaborate on how collections can meet these needs and offer design implications to enhance their capabilities.

The Role of the Collections in Supporting Sensemaking

To explain the collections' role in sensemaking, we used the model by Pirolli and Card [6], which divides the sensemaking process into 2 subcycles: the foraging loop and the sensemaking loop. On the basis of this model, collections can be described as a space for assembling relevant data about a topic, finding relationships between them, and storing outcomes from the sensemaking. In the foraging loop, patients gather relevant records to answer questions such as the following—"Is there a relationship between my weight and blood pressure?"—and save them in a collection, such as Weight vs. Blood Pressure. In the sensemaking loop, patients identify information relationships within the collection that they capture in notes, such as instances where there was co-occurrence of high blood pressure and high body mass. These notes help argue hypotheses such as the following: "My blood pressure is high when I'm

overweight." The outcome of this sensemaking process could be a comprehensive note for a clinical visit.

While, in their current form, collections respond to the needs of the sensemaking model by Pirolli and Card [6], improvements can be made to make this more efficient. This study revealed that medical records alone are not enough for reliable sensemaking. Adding PGD such as symptoms, measurements, outcomes, and everyday events is essential for creating comprehensive collections. The foraging loop can be made less laborious and time-consuming if there are additional visual cues, medical explanations, filtering capabilities, and automatic support to improve the relevance and reduce the effort of assembling collections. The sensemaking loop could be improved by adding more schematization capabilities such as annotating medical records and PGD to identify patterns later (eg, symptom triggers, medication effects, and correlations) and grouping records within collections, labeling those groups, and establishing group relationships with explanations (eg, linking "cholesterol lab results" with "food intake" as "food effects on cholesterol" in a High Blood Pressure collection).

Reliability of Collections

According to our study, there are 4 main factors that can influence the reliability of the collections: robust coverage of health issues, provision of PGD, grouping and linking of medical records and PGD within a collection, and verifying the contents of the collections. This reliability is related to collections' capability to aid in creating personalized but realistic health issue narratives, support self-management, and stimulate awareness and proactivity.

Robust Coverage of Health Issues: Relevance Assessment

Collections should ensure that patients can create collections for their most important health issues to support awareness and proactivity. Participants desired visual cues, explanations, and automatic support to determine which collections to create and what records to include.

While participants found the context-capturing data exploration using Sliding Tabs and Timeline convenient, they needed more to identify relevant records quickly. Future tools could incorporate Accordions that summarize record types, graph values over time, and highlight abnormal or extreme values. In addition, patients should be able to expand individual records to see explanations of clinical terms and clinical meaning interpretations. While there is existing work related to visualizing time series of EHR data [17-20] and automatic provision of short explanations [31], this study shows the need for combining them in a new way to support relevance assessment for a novel purpose—constructing collections. Finally, patients should be able to order and filter records by attribute for quicker browsing and bulk addition to a collection.

Automatic support should also be provided for creating and building collections to save time and ensure robust, reliable coverage of health issues. Our previous work highlighted the need for automating collections [13], and this study highlights a clear preference for automatically grouping medical records by clinical meaning—whether *thematically* or *guided by patient input*. Thematic collections would be those that tie records together based on conditions (bronchitis or diabetes) and procedures (stent placement or appendix removal) or with respect to organs (heart or kidneys) or organ systems (cardiovascular or renal). Alternatively, patients could specify a seed by setting parameters such as the collection's name, tags, or initial records. The system can offer candidate records to include or delete with confidence scores and explanations. Patients could then refine these system-generated collections by adding or removing records, PGD, and tags.

Addressing automatic support for collections may be challenging due to subtle relationships between medical records [32-34]. However, starting with easier constructs such as *time stamps* (eg, medical records from the same day, week, or month), *FHIR links* (eg, medical records from the same encounter or physician), *abnormal values* based on well-established clinical guidelines (eg, high or low blood pressure or cholesterol), *test findings* (eg, positive and negative), and *patient tags* (eg, triggers or pivotal events) is a feasible approach.

Provision of PGD: Sensemaking Data Completeness

Collections should include PGD to improve data completeness for sensemaking. Previous research has suggested that maintaining consistent PGD logging over time is difficult [35]. However, this should not be considered a barrier or a prerequisite for the collections' success. Patients' motivation and preferences for PGD logging intensity vary based on their disease self-management state [36]. When setting goals and learning strategies, patients prefer meticulous data collection. Once goals and strategies are in place, logging intensity decreases. In addition, if physicians require PGD logging for treatment planning, patients are motivated to engage in it [36,37].

Thus, collections should enable flexible and efficient PGD logging. Disease-specific contexts such as irritable bowel syndrome [38], diabetes [7], and migraine [39] have explored health sensemaking without focusing on diverse data types. This contrasts with patients' desire to log PGD for various medical issues within a single application [40,41] using a universal logging model for different observations [42,43]. To address this issue, we propose a straightforward workflow where patients initiate free-text entries and use tags to specify the type, quality, or other details. This mechanism allows for quick data capture and embellishment at convenient times.

Tags can classify PGD as *clinical observations*, *everyday life events*, or *notes*. Further specification can be added using tags such as *symptom*, *measurement*, *treatment*, and *outcome* for observations; *meal*, *exercise*, *meeting*, and *deadline* for life events; and *context*, *personal note*, and *visit note* for notes. Additional tags such as *absent*, *normal*, *high*, *low*, *extreme*, *improvement*, *deterioration*, *pivotal event*, *trigger*, or *relaxer* can be used for further detail that captures the quality and importance of the logged data. In addition to these system-offered tags, patients can also create their own custom tags for better personalization.

Grouping and Linking Within Collections: Schematization Capabilities

Patients need to connect medical records and PGD within collections for easier sensemaking. Future tools should add structure by enabling record grouping and linking of groups or individual records. This helps highlight important subsets of records and trace major conclusions as collections grow.

We recommend using a simple yet powerful tagging concept. Records sharing the same tag can form a group, whereas links between groups can be specified using related tags. The same mechanism can link individual records with other records or groups, providing nuanced sensemaking. This approach aligns well with the proposed PGD tagging model that can be applied to medical records as well.

Collection Verification

Collections represent the patient's personalized perception of their health and issues. As such, they should undergo occasional verification by the patient's physician for safe decision-making and action taking. While collection verification may add to the physician's workload, it can inspire and enable patients to manage health issues more independently. That said, patients should consider the physician's workload before requesting verification [44].

Future tools could allow for the conversion of a collection into a well–laid-out PDF document capturing all its contents. This document can be printed for review during a clinical visit or shared as a PDF attachment in the patient portal for verification at the physician's convenience.

Taking Actions Based on Collections

While annotating PGD is known to aid learning and disease self-management [7,45], our findings reveal that annotations can also enhance EHR data, creating synergy with PGD. Participants expressed a desire to annotate their data for various purposes: identifying triggers to avoid or encourage certain behaviors, marking pivotal events to remind them of shifts in health attitudes and management, and labeling outcomes as desired or undesired to evaluate treatments and strategies. These capabilities can be easily implemented using the previously elaborated tag-based design for linking records.

To increase the awareness and prioritization of collections, we previously proposed *collection descriptors* such as topic, urgency, currency, and sentiment [13]. Participants found value in these descriptors but expressed a need for additional ones that can be classified as *patient specified* and *data driven*. Future tools may include patient-specified descriptors for *clarity* (eg, is the diagnosis clear?), *stability* (eg, is the treatment working consistently?), *severity* (eg, is there a significant medical risk?), and *progress markers* (eg, is the issue substantially resolved?). Data-driven descriptors could be derived from the collection data, indicating the *time span* (from the oldest to the latest record) or listing the *physicians involved* (the providers and physicians the records came from). Both types of descriptors should be optional for patients to use as needed.

Providing an inner structure, enabling annotations, and describing collections can improve information access and expedite decision-making. Powerful search engines can use these metadata to allow patients direct and easy access not only to individual collections but also to their specific contents.

While these features can enhance collections' actionability, it is important to note that collections are not meant for making independent clinical decisions by patients. Collections should be verified by a physician to serve as reliable tools for sensemaking and health self-management. However, collections can always be invaluable tools for patients to understand their health; organize thoughts, hypotheses, and insights; and communicate effectively with their physicians.

Collection Use for Patient-Provider Communication

As observed in our findings, patients can use collections to prepare for a clinical visit by devising checklists and organizing thoughts supported by evidence. During the visit, collections can be used for *note taking* and, afterward, for *recording reminders* and *follow-up actions*. These uses indicate how collections can start addressing known challenges during clinical visits, such as problem presentation [46], information retention [47], setting common ground, aligning goals, and understanding instructions [48]. To effectively tackle such challenges, the Collections feature should support richer note capturing and collaborative data analysis in a collocated setting [49,50].

Future improvements in capturing PGD could make collections more appealing to physicians. For physicians, PGD play a crucial role in understanding the boundaries and context for accurate diagnosis and optimal treatment [51]. However, physicians often face problems with PGD, such as incomplete data, inconsistent data structures, and insufficient time for reviewing due to poor organization [52]. These issues arise because patients use disjointed platforms to log their data, lacking consistent models for logging different types of data [52,53], and face challenges in efficiently using these platforms [42,52]. Collections can help overcome these issues by providing a single platform for logging PGD for various health issues in a simple, universal way that allows for flexibility, organization, and standardization.

An alternative approach to enhancing collections as a communication tool and fostering physician collaboration in their creation and verification is to introduce them as a *shared resource* similar to Google Docs [54,55]. While this may seem unconventional, it builds on the principles of OpenNotes [56]. OpenNotes provides access to and transparency regarding clinical notes, enabling patients to improve their treatment and EHR data quality by taking an active role in detecting errors, raising concerns, asking questions, or seeking clarifications [56,57].

Similar to OpenNotes, *shared collections* would follow the principle of asynchronous communication and transparency. However, shared collections could eliminate the expressiveness constraints and lack of efficient ways to provide granular and tailored context observed in existing messaging systems [58,59]. In addition, shared collections would enable direct editing of underlying data in collaborative ways, minimizing communication overhead.

Moreover, shared collections would introduce a new communication channel between patients and providers outside the traditional patient portal. Synchronizing the digital traces of care in collections with the provider's EHR system to avoid discrepancies and legal issues should be a top consideration in future design iterations of the shared collection concept.

A Glimpse Into the Future: Collections and Generative Artificial Intelligence

In the future, we should explore the potential of generative artificial intelligence (GenAI) models to support patient sensemaking through collections. Tools such as ChatGPT [60] and Med-PaLM [61], which have demonstrated substantial medical knowledge [62-64], can replace the need for custom-made machine learning algorithms for knowledge-intensive tasks.

In particular, GenAI tools can aid in automatic and iterative collection construction with explanations and guidance. They can analyze the data within collections for insights, including medical records, PGD, notes, and tags. In addition, GenAI can assist in composing case narratives and talking points for clinical encounters. By offering this level of automation, GenAI can help tackle the significant knowledge challenges while lowering the labor barrier for patients' sensemaking activities.

Using GenAI models, collection construction could rely on natural language instructions such as the following: "Group my EHR data by condition," "Find all records related to my bronchitis," or "Identify records that don't belong to this collection and those that are missing." GenAI models could also deliver context, explaining why certain records are included or excluded and providing educational material such as term definitions and clinical implications.

In addition, patients can issue commands for identifying relationships within their annotated data, such as "List all triggers for my headaches over the last year." Finally, they can ask for help in constructing case presentations for clinical visits (eg, "Based on my 'High Blood Pressure' collection notes, write a 100-word summary").

To be useful for sensemaking, GenAI tools do not need to achieve complete accuracy. While still striving for maximum reliability, their main value should come from providing an environment that enables and encourages patients to refine artificial intelligence–generated outputs. As such, the contribution of GenAI toward sensemaking would be evaluated on its ability to help the patient efficiently produce a satisfactory solution with minimal physician input.

Finally, existing approaches for supporting sensemaking through search and interactive visualizations should not be disregarded. Exploring the integration of GenAI, search, and visualization is a prudent strategy as different sensemaking tasks related to collection assembly, editing, and analysis may require diverse approaches based on complexity, patient skills, and artificial intelligence reliability.

Limitations

This study has several limitations. First, the cohort skewed younger, likely due to recruitment via Craigslist (less popular with older adults) and the complexity of the remote setup. Second, participants used data from a fictitious patient, which may have reduced their motivation to learn the app and their ability to suggest real-life use cases. Third, participants had limited time to learn how to interact with collections, possibly affecting their perceptions of usability and utility. Future studies should have participants use their own data with automatic interaction logging. Despite these limitations, this study provided valuable insights into designing patient-facing sensemaking tools for organizing and augmenting EHR data.

Conclusions

Collections can potentially improve patient-centered care by involving patients more in decision-making and encouraging self-advocacy. Current assumptions often expect patients to have the necessary skills, tools, and motivation. We believe that collections can lower these barriers, encouraging patients to *increase engagement* with their health data, better *educate themselves*, and *communicate more effectively* with their care providers.

Our study suggests that EHR data can be better used and more useful for patients through *improved organization* and *annotation*. This approach can incentivize patients to engage more deeply with their EHR data, develop insights, and reflect on their experiences. Patients felt that this empowered their awareness, resourcefulness, and proactivity regarding health issues, making them more prepared and better informed for clinician interactions.

These findings support our premise that collections are a crucial step toward *patient empowerment* and *self-advocacy*. With appropriate improvements, collections can enhance patients' expertise by facilitating sensemaking activities and enabling insightful discussions with their physicians. First, collections motivate patients to construct health models based on their issues and ongoing problems. Second, patients gain medical education by actively participating in the evolution of collections through independent or system-assisted assembly and editing. Finally, patients acquire additional medical knowledge by engaging in meaningful discussions with their physicians and considering their feedback on collection verification.

Our study highlighted the importance of integrating PGD with EHR data. We envision a synergy in which patients use clinical data as a foundation, augmenting them with their observations, notes, and annotations to create personalized health narratives that support better health management and provider communication.

In the future, we should explore GenAI models to support patient sensemaking through collections. These models could help patients build collections, analyze the data within them, and produce health narratives more efficiently. Such enhancements may also reduce physicians' workload for verifying collection contents, leading to more focused, evidence-driven discussions during visits.

Promising ideas from this work should be advanced carefully, with gradual design improvements tested in real-life settings. Respecting existing clinical practices and workflows can facilitate quicker adoption and more significant changes in the

future. We believe that collections can revolutionize how patients interact with their medical records and communicate with their providers.

Acknowledgments

This work was funded by Scripps Research through the Technology to Empower Changes in Health Network Participant Technologies Center award (grant 5U24OD023176).

Conflicts of Interest

NG is a cofounder and equity owner of datavisyn.

Multimedia Appendix 1
Demographic and participant characterization questions and the evaluation tasks for the 2 study sessions—session 1 and session 2.
[DOCX File , 25 KB - humanfactors_v12i1e50331_app1.docx]

References

1. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):e240-e251. [doi: 10.1016/s1470-2045(20)30021-8]
2. Kwame A, Petrucka PM. A literature-based study of patient-centered care and communication in nurse-patient interactions: barriers, facilitators, and the way forward. BMC Nurs 2021 Sep 03;20(1):158 [FREE Full text] [doi: 10.1186/s12912-021-00684-2] [Medline: 34479560]
3. New applicability dates included in ONC interim final rule. The Office of the National Coordinator for Health Information Technology. URL: https://tinyurl.com/bdhxedu2 [accessed 2025-01-14]
4. ONC's cures act final rule. National Coordinator for Health Information Technology. URL: https://tinyurl.com/3upcepm4 [accessed 2025-01-14]
5. Naumer C, Fisher KE, Dervin B. SenseMaking: a methodological perspective. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI '08; April 5-10, 2008; Florence, Italy.
6. Pirolli P, Card SK. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: Proceedings of International Conference on Intelligence Analysis. 2005 Presented at: ICIA 2005; May 2-6, 2005; McLean, VA. [doi: 10.1145/1358628.1358972]
7. Raj S, Lee JM, Garrity A, Newman MW. Clinical data in context: towards sensemaking tools for interpreting personal health data. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-20. [doi: 10.1145/3314409]
8. Healthcare. Apple Store. URL: https://www.apple.com/healthcare/health-records/ [accessed 2024-09-08]
9. iBlueButton homepage. iBlueButton. URL: https://www.ibluebutton.com/ [accessed 2024-09-08]
10. OneRecord homepage. OneRecord. URL: https://onerecord.com [accessed 2024-09-08]
11. 1upHealth homepage. 1upHealth. URL: https://1up.health/product/patient [accessed 2024-09-08]
12. Nakikj D, Kreda D, Gehlenborg N. New ways for patients to make sense of their electronic health record data using the discovery web application: think-aloud evaluation study. JMIR Form Res 2023 Apr 03;7:e41346 [FREE Full text] [doi: 10.2196/41346] [Medline: 37010887]
13. Nakikj D, Kreda D, Gehlenborg N. 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. JMIR Hum Factors 2023 Aug 21;10:e41552 [FREE Full text] [doi: 10.2196/41552] [Medline: 37603400]
14. Zhang M, Ehrmann D, Mazwi M, Eytan D, Ghassemi M, Chevalier F. Get to the point! Problem-based curated data views to augment care for critically ill patients. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5,2022; New Orleans, LA. [doi: 10.1145/3491102.3501887]
15. Klein G, Phillips JK, Rall EL, Peluso DA. A data-frame theory of sensemaking. In: Hoffman RR, editor. Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making. Mahwah, NJ: Lawrence Erlbaum Associates Publishers; 2007.
16. Choe EK, Lee B, Zhu H, Riche NH, Baur D. Understanding self-reflection: how people reflect on personal data through visual data exploration. In: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017 Presented at: PervasiveHealth '17; May 23-26, 2017; Barcelona, Spain. [doi: 10.1145/3154862.3154881]
17. Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. LifeLines: using visualization to enhance navigation and analysis of patient records. Proc AMIA Symp 1998:76-80 [FREE Full text] [Medline: 9929185]
18. Wang TD, Plaisant C, Quinn AJ, Stanchak R, Murphy S, Shneiderman B. Aligning temporal data by sentinel events: discovering patterns in electronic health records. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI '08; April 5-10, 2008; Florence, Italy. [doi: 10.1145/1357054.1357129]

19. Rind A, Wang TD, Aigner W, Miksch S, Wongsuphasawat K, Plaisant C. Interactive Information Visualization to Explore and Query Electronic Health Records. Norwell, MA: Now Publishers Inc; 2013.
20. Gotz D, Borland D. Data-driven healthcare: challenges and opportunities for interactive visualization. IEEE Comput Graph Appl 2016;36(3):90-96. [doi: 10.1109/MCG.2016.59] [Medline: 28113160]
21. Welcome to FHIR®. Health Level 7 Fast Healthcare Interoperability Resources. URL: https://www.hl7.org/fhir/ [accessed 2024-09-07]
22. SMART health IT homepage. SMART Health IT. URL: https://smarthealthit.org/ [accessed 2024-09-01]
23. Martinez W, Threatt AL, Rosenbloom ST, Wallston KA, Hickson GB, Elasy TA. A patient-facing diabetes dashboard embedded in a patient web portal: design sprint and usability testing. JMIR Hum Factors 2018 Sep 24;5(3):e26 [FREE Full text] [doi: 10.2196/humanfactors.9569] [Medline: 30249579]
24. Faulkner L. Beyond the five-user assumption: benefits of increased sample sizes in usability testing. Behav Res Methods Instrum Comput 2003 Aug;35(3):379-383. [doi: 10.3758/bf03195514] [Medline: 14587545]
25. Haggstrom DA, Saleem JJ, Russ AL, Jones J, Russell SA, Chumbler NR. Lessons learned from usability testing of the VA's personal health record. J Am Med Inform Assoc 2011 Dec;18 Suppl 1(Suppl 1):i13-i17 [FREE Full text] [doi: 10.1136/amiajnl-2010-000082] [Medline: 21984604]
26. Yen PY, Walker DM, Smith JM, Zhou MP, Menser TL, McAlearney AS. Usability evaluation of a commercial inpatient portal. Int J Med Inform 2018 Feb;110:10-18. [doi: 10.1016/j.ijmedinf.2017.11.007] [Medline: 29331248]
27. Developer. Apple Store. URL: https://developer.apple.com/testflight/ [accessed 2024-09-16]
28. AirServer connect homepage. AirServer Connect. URL: https://www.airserver.com/ [accessed 2024-09-16]
29. Alhadreti O, Mayhew P. Rethinking thinking aloud: a comparison of three think-aloud protocols. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC. [doi: 10.1145/3173574.3173618]
30. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597. [doi: 10.1080/2159676x.2019.1628806]
31. Wilcox L, Morris D, Tan D, Gatewood J, Horvitz E. Characterizing patient-friendly "micro-explanations"of medical events. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI '11; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1978942.1978948]
32. Hirsch JS, Tanenbaum JS, Lipsky Gorman S, Liu C, Schmitz E, Hashorva D, et al. HARVEST, a longitudinal patient record summarizer. J Am Med Inform Assoc 2015 Mar;22(2):263-274 [FREE Full text] [doi: 10.1136/amiajnl-2014-002945] [Medline: 25352564]
33. Khedr AM, Aghbari ZA, Ali AA, Eljamil M. An efficient association rule mining from distributed medical databases for predicting heart diseases. IEEE Access 2021 Jan 19;9:15320-15333. [doi: 10.1109/ACCESS.2021.3052799]
34. Sánchez-de-Madariaga R, Martinez-Romo J, Escribano JM, Araujo L. Semi-supervised incremental learning with few examples for discovering medical association rules. BMC Med Inform Decis Mak 2022 Jan 24;22(1):20 [FREE Full text] [doi: 10.1186/s12911-022-01755-3] [Medline: 35073885]
35. Clawson J, Pater JA, Miller AD, Mynatt ED, Mamykina L. No longer wearing: investigating the abandonment of personal health-tracking technologies on craigslist. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Presented at: UbiComp '15; September 7-11, 2015; Osaka, Japan. [doi: 10.1145/2750858.2807554]
36. Li I, Dey AK, Forlizzi J. Understanding my data, myself:supporting self-reflection with ubicomp technologies. In: Proceedings of the 13th international conference on Ubiquitous computing. 2011 Presented at: UbiComp '11; September 17-21, 2011; Beijing, China. [doi: 10.1145/2030112.2030166]
37. Andersen T, Bjørn P, Kensing F, Moll J. Designing for collaborative interpretation in telemonitoring: re-introducing patients as diagnostic agents. Int J Med Inform 2011 Aug;80(8):e112-e126. [doi: 10.1016/j.ijmedinf.2010.09.010] [Medline: 21067968]
38. Karkar R, Schroeder J, Epstein DA, Pina LR, Scofield J, Fogarty J, et al. TummyTrials: a feasibility study of using self-experimentation to detect individualized food triggers. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17; May 6-11, 2017; Denver, CO. [doi: 10.1145/3025453.3025480]
39. Schroeder J, Chung CF, Epstein D, Karkar R, Parsons A, Murinova N, et al. Examining self-tracking by people with migraine: goals, needs, and opportunities in a chronic health condition. In: Proceedings of the 2018 Designing Interactive Systems Conference. 2018 Presented at: DIS '18; June 9-13, 2018; Hong Kong, China. [doi: 10.1145/3196709.3196738]
40. Tollmar K, Bentley F, Viedma C. Mobile health mashups: making sense of multiple streams of wellbeing and contextual data for presentation on a mobile device. In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare. 2012 Presented at: PervasiveHealth 2012; May 21-24, 2012; San Diego, CA. [doi: 10.4108/icst.pervasivehealth.2012.248698]
41. Hsieh CK, Tangmunarunkit H, Alquaddoomi F, Jenkins J, Kang J, Ketcham C, et al. Lifestreams: a modular sense-making toolset for identifying important patterns from everyday life. In: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems. 2013 Presented at: SenSys '13; November 11-15, 2013; Rome, Italy. [doi: 10.1145/2517351.2517368]

42. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. "You get reminded you're a sick person": personal data tracking and patients with multiple chronic conditions. J Med Internet Res 2015 Aug 19;17(8):e202 [FREE Full text] [doi: 10.2196/jmir.4209] [Medline: 26290186]
43. Jones SL, Hue W, Kelly RM, Barnett R, Henderson V, Sengupta R. Determinants of longitudinal adherence in smartphone-based self-tracking for chronic health conditions: evidence from axial spondyloarthritis. Proc ACM Interact Mob Wearable Ubiquitous Technol 2021 Mar 30;5(1):1-24. [doi: 10.1145/3448093]
44. North F, Luhman KE, Mallmann EA, Mallmann TJ, Tulledge-Scheitel SM, North EJ, et al. A retrospective analysis of provider-to-patient secure messages: how much are they increasing, who is doing the work, and is the work happening after hours? JMIR Med Inform 2020 Jul 08;8(7):e16521 [FREE Full text] [doi: 10.2196/16521] [Medline: 32673238]
45. Mamykina L, Mynatt E, Davidson P, Greenblatt D. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI '08; April 5-10, 2008; Florence, Italy. [doi: 10.1145/1357054.1357131]
46. Robinson JD, Heritage J. The structure of patients' presenting concerns: the completion relevance of current symptoms. Soc Sci Med 2005 Jul;61(2):481-493. [doi: 10.1016/j.socscimed.2004.12.004] [Medline: 15893061]
47. Unruh KT, Skeels M, Civan-Hartzler A, Pratt W. Transforming clinic environments into information workspaces for patients. Proc SIGCHI Conf Hum Factor Comput Syst 2010 Apr;2010:183-192 [FREE Full text] [doi: 10.1145/1753326.1753354] [Medline: 21654895]
48. Raj S, Newman MW, Lee JM, Ackerman MS. Understanding individual and collaborative problem-solving with patient-generated data: challenges and opportunities. Proc ACM Hum Comput Interact 2017 Dec 06;1(CSCW):1-18. [doi: 10.1145/3134723]
49. Chung CF, Wang Q, Schroeder J, Cole A, Zia J, Fogarty J, et al. Identifying and planning for individualized change: patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. Proc ACM Interact Mob Wearable Ubiquitous Technol 2019 Mar 29;3(1):1-27 [FREE Full text] [doi: 10.1145/3314394] [Medline: 31080941]
50. Schroeder J, Hoffswell J, Chung CF, Fogarty J, Munson S, Zia J. Supporting patient-provider collaboration to identify individual triggers using food and symptom journals. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR URL: https://doi.org/10.1145/2998181.2998276 [doi: 10.1145/2998181.2998276]
51. Chung CF, Dew K, Cole A, Zia J, Fogarty J, Kientz JA, et al. Boundary negotiating artifacts in personal informatics: patient-provider collaboration with patient-generated data. In: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2016 Presented at: CSCW '16; February 27-March 2, 2016; San Francisco, CA URL: https://doi.org/10.1145/2818048.2819926 [doi: 10.1145/2818048.2819926]
52. West P, Van Kleek M, Giordano R, Weal MJ, Shadbolt N. Common barriers to the use of patient-generated data across clinical settings. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC URL: https://doi.org/10.1145/3173574.3174058 [doi: 10.1145/3173574.3174058]
53. Choe EK, Lee NB, Lee B, Pratt W, Kientz JA. Understanding quantified-selfers' practices in collecting and exploring personal data. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, ON URL: https://doi.org/10.1145/2556288.2557372 [doi: 10.1145/2556288.2557372]
54. Google Docs. Facebook. URL: https://www.facebook.com/GoogleDocs/ [accessed 2024-08-01]
55. Jung YW, Lim YK, Kim MS. Possibilities and limitations of online document tools for design collaboration: the case of Google Docs. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. 2017 Presented at: CSCW '17; February 25-March 1, 2017; Portland, OR URL: https://doi.org/10.1145/2998181.2998297 [doi: 10.1145/2998181.2998297]
56. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
57. Fossa AJ, Bell SK, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 01;25(9):1153-1159 [FREE Full text] [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
58. Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L. Characterizing patient-clinician communication in secure medical messages: retrospective study. J Med Internet Res 2022 Jan 11;24(1):e17273 [FREE Full text] [doi: 10.2196/17273] [Medline: 35014964]
59. Eschler J, Liu LS, Vizer LM, McClure JB, Lozano P, Pratt W, et al. Designing asynchronous communication tools for optimization of patient-clinician coordination. AMIA Annu Symp Proc 2015;2015:543-552 [FREE Full text] [Medline: 26958188]
60. Introducing ChatGPT. OpenAI. 2022 Nov 30. URL: https://openai.com/blog/chatgpt [accessed 2024-06-21]
61. Med-PaLM. Google. URL: https://sites.research.google/med-palm/ [accessed 2024-06-21]

62. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023 Feb 08;9:e45312 [FREE Full text] [doi: 10.2196/45312] [Medline: 36753318]
63. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb 9;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
64. Singhal K, Tu T, Gottweis J, Sayres R, Wulczyn E, Hou L, et al. Towards expert-level medical question answering with large language models. arXiv Preprint posted online on May 16, 2023 [FREE Full text] [doi: 10.48550/arXiv.2305.09617]

Abbreviations

EHR: electronic health record
FHIR: Fast Healthcare Interoperability Resources
GenAI: generative artificial intelligence
PGD: patient-generated data
RQ: research question

Edited by A Kushniruk; submitted 01.09.23; peer-reviewed by S Young, N Finn; comments to author 15.11.23; revised version received 05.02.24; accepted 25.11.24; published 03.02.25.

Please cite as:
Nakikj D, Kreda D, Luthria K, Gehlenborg N
Patient-Generated Collections for Organizing Electronic Health Record Data to Elevate Personal Meaning, Improve Actionability, and Support Patient–Health Care Provider Communication: Think-Aloud Evaluation Study
JMIR Hum Factors 2025;12:e50331
URL: https://humanfactors.jmir.org/2025/1/e50331
doi:10.2196/50331
PMID:

Original Paper

Mental Health Providers' Challenges and Solutions in Prescribing Over Telemedicine: Content Analysis of Semistructured Interviews

Julia Ivanova[1], PhD; Mollie R Cummins[1,2], RN, PhD; Hiral Soni[1], MS, MBA, PhD; Triton Ong[1], PhD; Brian E Bunnell[1,3], PhD; Esteban López[1], MD, MBA; Brandon M Welch[1,4], PhD

[1]Doxy.me Research, Doxy.me Inc, Charleston, SC, United States

[2]Department of Biomedical Informatics, College of Nursing and Spencer Fox Eccles School of Medicine, University of Utah, Salt Lake City, UT, United States

[3]Department of Psychiatry and Behavioral Neurosciences, Morsani College of Medicine, University of South Florida, Tampa, FL, United States

[4]Biomedical Informatics Center, Medical University of South Carolina, Charleston, SC, United States

Corresponding Author:
Julia Ivanova, PhD
Doxy.me Research
Doxy.me Inc
3rd Floor Suite 6 & 7
18 Broad Street
Charleston, SC, 29401
United States
Phone: 1 8444369963
Email: julia.ivanova@doxy.me

Abstract

Background: In response to the COVID-19 pandemic, the United States extended regulatory flexibilities to make telemedicine more accessible to providers and patients. Some of these flexibilities allowed providers to intake patients over telemedicine and prescribe certain scheduled medications without an in-person visit.

Objective: We aim to understand providers' parameters for their comfort in prescribing over telemedicine and report on solutions providers have adopted in response to potential barriers and challenges in prescribing via telemedicine.

Methods: As part of a larger mixed methods study between February and April 2024, we conducted 16 semistructured interviews with mental health providers who prescribe via telemedicine within the United States. We used the results of a web-based, cross-sectional survey to develop a codebook and support recruitment. We analyzed a subsection of the 16 interviews using content analysis to capture comfort, barriers, and workarounds in telemedicine prescribing. We reported codes by frequency and by provider.

Results: Participants were typically male (11/16, 69%), provided care mostly or completely over telemedicine (11/16, 69%), and were psychiatrists (8/16, 50%) or other physician (3/16, 19%). Providers' primary states (10/16, 62%) of practice included Oregon, Texas, New York, and California. The content analysis yielded a total of 234 codes, with three main codes—comfort (98/234, 41.9%), barriers or challenges (85/234, 36.3%), and workarounds or solutions (27/234, 11.5%)—and two subcodes—uncomfortable prescribing (30/98, 31%) and comfortable prescribing (68/98, 69%) over telemedicine. Participants reported being comfortable prescribing over telemedicine as long as they could meet their main parameters of working within their expertise, having access to needed patient health information, and being compliant with rules and regulations. Participants reported frustrations with e-prescription workflows and miscommunications with pharmacies. Solutions to ease frustrations and alleviate discomforts in prescribing over telemedicine included developing workflows to help patients complete laboratory tests and physical examinations and directly communicating with pharmacies.

Conclusions: By applying content analysis to the semistructured provider interviews, we found that physicians are comfortable prescribing via telemedicine when they feel they are practicing within their personal parameters for safety. While many providers experience frustrations such as miscommunication with pharmacies, these barriers appear to not prevent them from telemedicine prescribing. With expected changes in 2024 and 2025 to the US laws and regulations for telemedicine prescribing, we may see changes in provider comfort in prescribing.

(JMIR Hum Factors 2025;12:e65419) doi:10.2196/65419

KEYWORDS

telemedicine; telehealth; prescribe; prescription; drug; pharmacology; pharmacotherapy; pharmaceutical; medication; barrier; buprenorphine; mental health; digital health; informatics; qualitative analysis; content analysis; provider perspective; provider; experience; attitude; opinion; perception; perspective

Introduction

During the height of the COVID-19 pandemic, health care visits moved to telemedicine when possible to help stem the spread of infection [1]. In the United States, the public health emergency (PHE) and regulatory flexibilities were instrumental in streamlining this effort by waiving specific requirements, such as the providers' ability to see patients outside the states where they are licensed and to prescribe over telemedicine with fewer restrictions [2,3]. Due to these temporary changes, providers could prescribe via telemedicine less restrictively and determine what level of care they felt comfortable providing over telemedicine.

The US Drug Enforcement Administration extended prescribing flexibilities past the May 11, 2023, PHE deadline and is in the process of finalizing a rule regarding telemedicine prescribing [4,5]. There has been concern that waiving elements of the Ryan Haight Act, which previously required providers to meet with patients in person before prescribing certain controlled substances, would adversely impact patient outcomes [6,7]. The concurrent opioid epidemic has also provided a new level of scrutiny in the space of telemedicine and prescribing scheduled medication for the treatment of substance use disorder (SUD) [8-10]. As US providers await the promulgation of this final rule, evidence continues to grow regarding positive patient outcomes via mental health prescribing over telemedicine [11]. One Medicaid data study following over 90,000 patients showed remarkable positive outcomes from telemedicine initiation of buprenorphine treatment, including better odds of 90-day treatment retention than if initiation occurred in person [12]. In a scoping review from 2008 to March 18, 2021, reviewers found that telehealth technology in SUD treatment increased access and adherence to buprenorphine and generally showed higher patient satisfaction and comparability to in-person retention rates [11]. Especially within the space of buprenorphine prescription, evidence supports that telemedicine offers comparable treatment to in-person care and can be an effective method for increasing patient access to mental health treatments [8,11].

With the flexibilities, providers can now choose the extent to which they use telemedicine. One study reported psychiatrists felt telemedicine use in a hybrid scenario allowed for treatment adaptation based on client needs and access [13]. The study's participants emphasized their need to consider each patient's unique case in determining whether they could successfully leverage telemedicine to initiate care or ensure continued visits over time [13]. For example, one psychiatrist in the study noted the following:

> *When you have these patients who are at risk for falling out of care, if you don't offer them telemedicine, but [they] are also at risk for getting slightly suboptimal care when you do offer them telemedicine, it is a very case-by-case judgment call in terms of the risks and benefits of enabling the telemedicine.*

Such sentiments support the call for greater provider autonomy and options, as well as the focus on including telemedicine as part of the holistic health care model [13-15].

Though there is evidence supporting the safety of prescribing certain scheduled substances over telemedicine (even without an initial in-person visit) [8,11,12], there is a need for greater insight into mental health providers' experiences in prescribing scheduled medications during and after the PHE. A better understanding of providers' perspectives, experiences, and comfort in prescribing for their patients over telemedicine would build a more accurate understanding of what providers view as actual challenges and limitations. This knowledge would add to the discourse regarding updates in policy and guidelines, especially necessary for policy makers as they deliberate over proposed rules and regulations [16]. To the best of our knowledge, no previous studies have focused on mental health providers' experiences regarding general prescribing—including scheduled medications—over telemedicine. In early 2024, we conducted a mixed methods study that included a web-based, cross-sectional survey and semistructured interviews of mental health providers who prescribe via telemedicine. The survey aimed to assess provider perceptions, experiences, comfort, and perceived patient safety in prescribing medications over telemedicine [17]. The interviews aimed to get a better understanding of providers' general telemedicine use, challenges and solutions specific to prescribing over telemedicine, and providers' knowledge and experiences regarding compliance with prescribing over telemedicine. In this study, our objectives were (1) to understand better providers' parameters for safe prescribing over telemedicine and (2) to report practices or workflows providers have adopted to accomplish safe prescribing via telemedicine. The results of the study will help stakeholders reassess long-standing policies—made at the beginning of these technological advances—that may now be causing unintended harm as the technology has improved and become more accessible to laypeople.

Methods

Study Design

We performed exploratory content analysis of semistructured interviews with tele–mental health care providers as part of a larger, mixed methods study. We used an explanatory sequential design, building upon a previously reported, web-based, cross-sectional survey [17].

Interview Development

After an initial literature review, we developed the semistructured interview guide with consultation and pretesting from 5 providers. We estimated that interviews would take 60 minutes. The interview attempts to capture context regarding what factors affect providers' views and comfort in prescribing medications via different scenarios and combinations of in-person and telemedicine visits. The interview guide covers the topics of (1) the provider's practice, such as specialty, telemedicine use, patients, and general difficulties with telemedicine; (2) challenges and solutions concerning prescribing over telemedicine; and (3) knowledge pertaining to laws and regulations of prescribing to telemedicine (see Multimedia Appendix 1 for the interview guide). This study focuses specifically on answers to the challenges and solutions concerning prescribing over telemedicine (Multimedia Appendix 1).

Participants and Recruitment

We recruited participants for the semistructured interviews from among participants in a research survey. The survey sampling, recruitment, and results are reported elsewhere [17]. In brief, we recruited a sample of US tele–mental health care providers from a telemedicine research panel, TelehealthEngage [18]. A total of 115 participants fully completed the survey. Participants were predominantly White (83/115, 72.2%), non-Hispanic or Latino (97/115, 84.3%), female (66/115, 57.4%) providers who were physicians (69/115, 60%) seeing their clients most to all of the time (50%-100% of visits) over telemedicine (82/115, 71.3%) [17].

Participants who agreed to be contacted at the end of the survey were invited to participate in follow-up interviews. Of the 115 completed surveys, 59 (51.3%) participants showed interest in being interviewed. Research team member JI (a White, immigrant female individual with a background in medical anthropology, psychology, and biomedical informatics) sent out interview invitations and completed interviews between February and April 2024. We aimed to complete 20 interviews or until we reached content saturation [19].

Data Analysis

One author (JI) performed an exploratory content analysis of 6 interviews. The codebook for content analysis, consisting of three main codes and two subcodes, was specified on the basis of prior survey findings (Table 1) [17]. The codebook was developed to test our expectations that providers will feel comfortable when their personal parameters for patient safety have been met and when the telemedicine visit is within the scope of provider practice and expertise. We used content analysis of the interviews to identify codes related to personal parameters for comfort in prescribing over telemedicine, named barriers and challenges of telemedicine prescribing, and the types of workarounds providers have found useful within those situations [19].

Table 1. Demographics as reported by providers during the interview[a].

Demographics for content analysis	Value, n (%)
Sex (n=16)	
Male	11 (69)
Female	5 (31)
How often telemedicine is used (n=16)	
None to seldom	1 (6)
Fairly often (about half the time)	4 (25)
Mostly or completely	11 (69)
Provider type (n=16)	
Psychiatry	8 (50)
Family medicine or general practice physician	3 (19)
Nurse practitioner	3 (19)
Physician assistant	2 (13)
Type of practice (n=21)	
Private practice	12 (57)
Statewide health system	4 (19)
Community health center	2 (10)
Solo practice	1 (5)
City-run program	1 (5)
Outpatient rehabilitation center	1 (5)

[a]Some demographic values sum to more than 16 as some providers worked in multiple practices.

We transcribed and deidentified completed interviews using Dovetail software (Dovetail). We confirmed transcripts for accuracy prior to deleting video data of interviews and uploaded transcripts to MAXQDA 2024 software (MAXQDA) for analysis. A single author (JI) coded the interviews over three iterations using the previously developed codebook, and coauthor HS (an Asian, immigrant female individual with a background in biomedical informatics and human factors research in behavioral health) reviewed the codes and codebook for validity and accuracy. We resolved discrepancies through consensus to ensure consistency and accuracy in the qualitative method [20,21]. The unit of analysis was meaningful phrases. Although we calculated the total frequency of code instances, we primarily focused on the frequency of codes by individual participants [19]. Additionally, contextual subtopics within the main reported codes were often double coded—as participants provided multiple examples, explanations, and reasoning—resulting in percentages over 100. We reported participant characteristics in the aggregate to ensure participants' confidentiality. For context, we included participants' numbers (1 through 16) after each quote.

Ethical Considerations

This study was reviewed and approved by the BRANY Institutional Review Board (IRB00010793). We attained verbal informed consent for the interviews in addition to the written informed consent from the survey protocol. Participants had the opportunity to opt out of our research. Participants were compensated with a US $75 e-gift card for their time. While there is always a risk of selection and other compensation-related biases, the amount was deemed appropriate for US mental health professionals who were the focus of recruitment. We use deidentified data for the purposes of this research.

Results

Demographics

We completed 16 semistructured interviews between February 22 and April 26, 2024. There were 11 (69%) male and 5 (31%) female participants (see Table 1 for demographics). Participants primarily practiced in the following states: Oregon (n=3, 19%), Texas (n=3, 19%), New York (n=2, 12%), Illinois (n=2, 12%), California (n=1, 6%), Florida (n=1, 6%), Michigan (n=1, 6%), Washington (n=1, 6%), Indiana (n=1, 6%), and Tennessee (n=1, 6%). The majority (n=11, 69%) provided care mostly or completely over telemedicine (75%-100% of visits), 4 (25%) providers reported an even mix (40%-60% of telemedicine and in-person visits) in their hybrid practice often depending on the weather, and 1 (6%) provider reported almost complete in-person care except for emergencies. Providers included psychiatrists (n=8, 50%), family medicine or general practice physicians (n=3, 19%), nurse practitioners (n=3, 19%), or physician assistants (n=2, 12%). Providers reported working in private practice (n=12, 75%), statewide health systems or clinics (n=4, 25%), community health centers (n=2, 12%), a solo practice (n=1, 6%), a city-run program (n=1, 6%), and an outpatient rehabilitation center (n=1, 6%). Some providers worked in multiple positions or practice types.

Content Analysis

A total of 234 code instances were identified, corresponding to three main codes and two subcodes. Table 2 shows the frequency of code instances corresponding to each main code and subcode. Frequencies of instances of *generally uncomfortable* and *generally comfortable* subcodes add up to the total frequencies of the *comfort* main code instances.

Of the 98 (41.9%) out of 234 total code instances under *comfort* in prescribing medications over telemedicine, 69% (68/98 code instances) were related to describing positive comfort in prescribing from all 16 providers. However, the majority of providers (10/16, 62%) also reported varying discomfort (30/98, 31% code instances) in prescribing over telemedicine even though it was less frequently mentioned.

Providers reported (85/234, 36.3%) varied barriers and challenges (85 code instances, 14 providers), providing information regarding problems that both providers and their patients face when prescribing medications over telemedicine. Providers reported fewer (27/234, 11.5%) workarounds and solutions (27 code instances, 11 providers).

Table 2. Content analysis results with the frequency of instances by main codes and subcodes[a].

Name	Frequency of code instances, n	Definition	Example
Comfort (main code)	98	Any direct discussion regarding feelings of comfort or ease in prescribing over telemedicine	"I just don't wanna like, yeah, miss something. And like, actually probably more so with the Benzos than with, than with stimulants because I feel like the stimulants you can: see if people are kind of like … when people are on Benzos they might be more drowsy, that you're just not picking up with, on the, the, the, the virtual…" [Participant 1]
Generally uncomfortable (subcode)	30	Any direct discussion regarding feelings of discomfort or unease in prescribing over telemedicine	"I feel less comfortable with over telemedicine because sometimes just like the cadence of a conversation is hard. Sometimes if there's a lag in telehealth or like whatever it may be. As for prescriptions, I would say, like your level twos and level threes are ones that I would not consider prescribing unless I'd at least seen them once in person…" [Participant 3]
Generally comfortable (subcode)	68	Any direct discussion regarding positive feelings of comfort or ease in prescribing over telemedicine	"And I would say that it's really about the same, the actual prescribing is pretty darn smooth. So, I feel like the mechanics of the prescribing are the same. I do not prescribe differently because I, I've got a sense, I don't know this person as well because I'm not me. No, I feel like I know my patients really well. Certainly, before I'm gonna be prescribing, I know what the story is in my mind. I may be wrong, but I've got that level of certainty. So, I don't think the telehealth platform impacts my decision-making or the mechanics around prescribing significantly." [Participant 6]
Barriers and challenges (main code)	85	Any direct discussion where a provider mentions a problematic situation that affects their decision-making to prescribe over telemedicine or affects a patient's ability to receive treatment	"But I would say, yes, it's difficult to feel comfortable, directing someone in another state to something that they may need without really knowing where that resource, who that resource is." [Participant 7]
Workarounds and solutions (main code)	27	Any direct discussion where the provider describes a way they adapt their process or decision-making due to a barrier or challenge	"So, a lot of times I give them information on how to find a particular how to vet another practitioner, but that's about the best [I can do]." [Participant 7]

[a]Main code and subcode definitions were developed based on survey analysis.

Parameters for Comfort in Prescribing Over Telemedicine

Overview

Providers brought up three main parameters as they contextualized their comfort or discomfort with prescribing medications over telemedicine: (1) limits of telemedicine and personal expertise, (2) knowledge of patients and access to their relevant health information, and (3) liability concerns.

Limits of Telemedicine and Personal Expertise

Five (31%) of the 16 providers established that they felt comfortable when the interaction stayed within their perceived limits of telemedicine and occurred within their realm of expertise. One provider pointed out that these parameters essentially establish criteria for a successful telemedicine visit:

> *So the direct answer to the question is I feel very comfortable [prescribing over telemedicine]. But I think the reason that I feel very comfortable has to do with the patients: who I see and who I won't see on telemedicine. So again, doing this is like a philosophical thing. I'm a big believer in getting people to the right place. And a lot of times that's me and a lot of times that's not me when it comes to mental health and telehealth, I try to, there are things that I, I don't personally feel comfortable or think would be appropriate to, to treat. So in general, if someone is like in a manic state, if they're psychotic, if they're actively suicidal, like suicidal with a plan intending to act on it, that sort of stuff, I won't see them through telehealth or if I have a patient who starts to experience those things, I will refer them for a higher level of care and we'll be pretty insistent on it. Not that I'm not gonna see them anymore, but it's just I want to get you to the best place.* [Participant 13]

Regarding the parameter of experience and expertise, 3 (19%) providers reported feeling uncomfortable prescribing over telemedicine when they thought they did not have the expertise to treat a particular diagnosis. One provider described feeling nervous about such a scenario:

> *I don't have anybody that I'm treating for narcotic abuse or opiate stuff. So I, not that I wouldn't, but I don't have anybody and I don't have a lot of experience with that. So that would make me nervous, especially over telemed. I just because I don't do it much in person either.* [Participant 9]

In this case, the provider notes they would feel uncomfortable treating such a patient in person and that seeing such a case over telemedicine would increase that feeling.

Knowledge of Patients and Access to Their Relevant Health Information

This parameter encompassed several topics including working with established patients and accessing needed patient health information such as labs or physical exams. A total of 4 (25%) providers identified knowledge of a patient's history—especially in regard to them being an established patient—as a crucial element of feeling comfortable prescribing for them.

> *So, certain medication that requires good monitoring for that level that may make me not want to prescribe medications on this platform. But I don't think it has anything to do with virtual platforms, any providers or psychiatrists would do so in, in person also. So I'm not, I'm trying to think if, unless the only time I don't prescribe is if someone makes the first appointment and say I'm on this medicine and I need refill, I don't do that because that's not a good practice. It's like, OK, unless I have some data or previous records or not a very strong medication, then you can continue. But if someone comes with a specific request for medicines that are not appropriate, then I wouldn't consider that as an OK to do it as a virtual practice.* [Participant 10]

This provider noted that in the scenario provided, they would need more information about the patient before prescribing, regardless of whether the visit is in person or via telemedicine. Meanwhile, one of these providers mentioned the importance of generally knowing their patients and their health status as they determine whether telemedicine visits are sufficient for care:

> *...If you have somebody who's a fragile diabetic, you need to have a different level of observation for that person. But if you have a stable diabetic, then you can just continue to prescribe that whatever it might be…* [Participant 15]

Inherent in this explanation is that the provider felt they have sufficient knowledge regarding a patient's current and past health status and feeling personally able to diagnose and treat in the presenting situation.

A total of 4 (25%) providers commented on the importance of having established patients in a scenario to make providers feel more comfortable prescribing over telemedicine. One of these providers underlined the importance of having such a relationship, especially when seeing patients for SUD treatment, as they may be able to pick up on potential concerns requiring more oversight via telemedicine or in person:

> *I mean, people with substance use problems. Now that brings a whole host of challenges itself because of confidence in, in history, you know, and confidence of just general reporting ... So I have a special, I have a different kind of relationship with [specific patient with SUD]. I can tell when somebody's bullshitting. And I can call them pretty well on it. I rarely get blindsided because I've been doing this for a long time. So I come with enough experience, I think also that allows that to happen. But suffice to say that if I have an issue with someone there, I'll put more controls over it, and we'll go to task if needed. If they need to come in, I need to see them in person...* [Participant 15]

In this case, the provider leveraged not only their years of experience in the specific field (expertise parameter) and having known the patient for a long time but also their hybrid practice, as they can have the patient come for an in-person visit.

A total of 3 (19%) providers mentioned they feel comfortable prescribing over telemedicine because they have such an option. In addition to this, 4 (25%) providers also pointed out that having a way to check on their patient's safety, whether it be having local resources for the patient or being able to view a Prescription Drug Monitoring Program, influences their comfort positively. One of these 4 providers described that knowing they can get patient health information quickly and seamlessly made them comfortable prescribing for an individual:

> *If they have established care, either someone who can update me on their vitals or regular monthly visits and have some kind of formal evaluation for me to backtrack the diagnosis. And then there is a way to order urine drug screens to make sure they're not abusing any other drugs on the top of. So, if that makes it an easy flow for us, then, yeah. Yeah. I think that's not a big problem. It's just how to coordinate that care sometimes makes it harder...* [Participant 10]

Additionally, 4 (25%) providers stated the importance of developing and adhering to a protocol with their patients. Providers discussed this concept within the patient information parameter in everything from the formal quality of care processes to simply expecting to see a patient in person for all intakes. Adhering to protocols appears to be driven by the need for patient health status and history, as providers often discussed continuity of care and procurement of labs and exams.

> *So, I think patients who are, really complex, I feel less comfortable with prescribing over telemedicine because sometimes just, like, the cadence of a conversation is hard. Sometimes if there's a lag in telehealth or whatever it may be. As for prescriptions, I would say, your level two's and level three's are ones that I would not consider prescribing unless I'd at least seen them once in person. I know personally, our clinic does random drug tests on our patients...* [Participant 3]

Liability Concerns

A total of 5 (31%) providers noted that they maintain their comfort in prescribing over telemedicine by simply refusing to prescribe certain controlled substances or not accepting specific diagnoses in their practice, especially due to their wariness of laws and regulations. One of these providers explained that while they treat SUDs, they do not prescribe certain medications and are aware of prescribing barriers:

> *...I do see a lot of patients with substance use disorder...just the fact that it's [buprenorphine] more accessible now is amazing for these patients. I've heard there are still some barriers with methadone in particular...it hasn't been an issue for me because I don't prescribe that medication in particular.* [Participant 16]

Indeed, 2 (12%) providers indicated they avoid prescribing over telemedicine unless they feel they have to for the patient.

> *I generally refrain from prescribing anything scheduled over telemedicine ... just because of the, the insane liability with it ... the unbalanced liability, I would say that physicians shoulder in this industry.* [Participant 7]

Here, the provider also touched on the issue of compliance and liability inherent in this situation.

Barriers and Workarounds in Prescribing Over Telemedicine

Of the 85 (36.3%) out of the 234 total code instances that describe barriers (14 providers), 37 (44%) of the 85 instances specifically mentioned that the e-prescription platform and the pharmacies caused challenges in prescribing medicines or picking up medicines (13 providers). There were 17 (20%) instances of providers describing difficulties with the actual e-prescription platform they use.

> *But Epic, they started putting in all these like hard blocks like you need to click this before you can do what you're trying to...I think the way that they went about it sometimes, it was really like a hassle for the workflow when you add it all together.* [Participant 2]

Providers also often noted their patients' prescription order was not received or not fulfilled by the pharmacy due to shortages in medicines or misunderstandings in prescribing over telemedicine (20/85, 24%). One provider noted how such a situation acted as a barrier for patients receiving their necessary prescriptions:

> *But I had a pharmacy once, call me that said that...they weren't going to allow my patient to fill a pack of the prescription because I hadn't provided an estimated glomerular filtration rate or creatinine clearance with my prescription. But like that's because they now have the ability to prescribe that on their own. But I sent the prescription. I'm assuming the responsibility. I already checked that all you're doing is throwing up another barrier to this patient, getting their medication when you do that, which I get. You don't want to be audited or whatever, but you're not the one prescribing it. I am, you know what I mean? I just like the pendulum has swung so far the other way since the opioid epidemic that they're really trying to tighten the noose and on all this stuff, and I just think it's going to end up harming patients who are calling for their beta blockers or, you know, whatever.* [Participant 2]

Providers reported shortages of certain medicines, usually stimulants, in pharmacies, requiring patients to find pharmacies with stock and providers to resend prescriptions. Providers' complaints regarding the e-prescription platforms and pharmacies' inability to provide the medications ranged from workflow nuisances to concerns that their patients could not receive necessary medications promptly.

Other mentioned barriers (49/85, 58%) were related to general telemedicine concerns and issues such as patient health concerns (17/85, 20% codes; 11/16, 69% providers), rules and regulations (16/85, 19% codes; 7/16, 44% providers), and reimbursement concerns (9/85, 11% codes; 2/16, 13% providers). Patient safety concerns mainly touched on worries that patients are not having their follow-up appointments, exams, and labs in a timely manner, which can affect patient health outcomes, as well as the ability of providers to prescribe medications for them.

> *I will put a child on a medication and I'll say I need to have you call me in two or three days because this medicine doesn't take a long time to start working. And I need to know so we can make adjustments. And invariably 99% of the time they wait till the next month, scheduled appointment and we could have made some, you know, needed changes. And that was difficult even when we were in the office...* [Participant 11]

Concerns regarding rules and regulations more closely align with providers seeing patients across state borders. Some providers felt such regulations hindered their ability to treat patients but also created unnecessary frustrations.

> *So, yeah, it's interesting navigating all of that because, and honestly, I think it's just so stupid having state-dependent licensing system when they all use the same criteria.* [Participant 7]

While providers recognized the need for laws and regulations, their current struggles often placed them in positions having to choose between seeing their (often established) patients or complying with regulations.

> *It reminds me of another patient, he's in college in North Carolina and, he had really good, he had a psychiatrist through his school but then the school had canceled the contract with that [psychiatrist] and he's like, what now am I supposed to do? And like, he wasn't exactly, he wasn't unstable but he wasn't stable. Like, you know, in that gray zone and he's like, can't you see me?...And I would love to see you. I feel my heart goes out to you, but like, you're in North Carolina, like, I...I, I will see you all through Christmas break, summer break. But like, I don't know what to do. Like, if you're having a crisis, definitely call me like, and I will, you know, I will just override it. But like, if you're going there and, you know, like you need the help, like you gotta, I don't know, I don't know what to tell you. Like, I feel terrible if you.* [Participant 1]

Providers often reported that mobile patients such as college students are the ones most strongly affected by these types of laws and regulations.

Providers reported far fewer workarounds or solutions (27/234, 11.5% codes; 11/16, 69% providers) to mentioned barriers relating to prescribing over telemedicine. Direct responses to barriers and challenges included—as subtopics—streamlining how they receive patient health data (8/27, 30%), finding new pharmacies for patients to use (5/27, 19%), using reminders and new ways to connect with patients quickly (5/27, 19%), responding to regulatory pressures (3/27, 11%), and making individual adaptations in their practice (6/27, 22%; due to double coding of subtopics, percentages add up over 100) The most common solution providers reported was finding an easier way to receive laboratory tests, examinations, and vitals from their patients (8/27, 30% codes; 5/16, 31% providers). While providers were able to direct their patients to local laboratories or request that their patients invest in a blood pressure cuff, one provider noted their practice found certain tests to be beneficial for telemedicine visits:

> *Hm, we do drug screens online. So, we do saliva testing online and drug screens. And so that's if we don't always have to have people go pee somewhere or go to the lab, they can do those at home. So, I guess that's something that we found just as a, not really a workaround, well, kind of a workaround, and I guess to allow them to stay at home, but just to incorporate it as something that's a little bit more convenient.* [Participant 5]

Providers further noted the emphasis on streamlining patient processes by using novel ways to communicate with them regarding simple questions or requests. For example, one provider noted the following:

> *[Patients] can just text me a question...A few patients of mine have learned that they come and sit in the waiting room if they have a question. And so, I see them sitting in my waiting room, and sometimes I'm done. I'm like, "What's up?" And they say, "Yes, you know, hey, I need a refill.*" [Participant 12]

In this case, the provider sought to have secure, simple messaging with a patient. Other solutions related to streamlining processes with patients included actively talking with pharmacies in the patients' area (5/27, 19%), moving to a new e-prescription platform (1/27, 4%), and creating a protocol for helping their out-of-state patients find health resources (1/27, 4%).

Providers mentioned two differing approaches to handling the barriers created by laws and regulations in prescribing over telemedicine. Some (2/16, 13%) providers chose to unequivocally adopt the strictest state and federal regulations within their practice, and another explained that they would continue with a process that works for their patients but was admittedly in the gray zone legally.

Content analysis of barriers affecting provider comfort in prescribing over telemedicine identified 85 code instances over 16 interviews. With 27 code instances of workarounds and solutions directly relating to these noted barriers, the qualitative analysis identifies areas in telemedicine where progress still needs to be made.

Discussion

Principal Findings

By leveraging the results of our more extensive mixed methods study [17], we framed our content analysis to provide needed context for understanding mental health providers' comfort in prescribing over telemedicine. This qualitative study identified vital parameters that constitute comfort for mental health providers who prescribe via telemedicine. Our results show that providers feel most comfortable prescribing over telemedicine when they are practicing within its inherent limitations and their own professional expertise, with access to patients' relevant health information, and within clearly defined legal and regulatory confines. In addition, we identified related barriers that affect these parameters and the potential solutions that providers have used to alleviate these challenges. Content analysis revealed providers found the e-prescription platforms and the miscommunications with pharmacies receiving prescriptions as substantial sources of frustration and concern in the prescribing process. Ultimately, providers prescribe medications over telemedicine within the bounds of their perceived comfort parameters but would like to see changes in streamlining the prescribing workflow, assurance of timely patient examinations and laboratory tests, and clarity in the laws and regulations surrounding prescribing over telemedicine.

Overall, even while providers reported feeling comfortable prescribing over telemedicine, they were aware their comfort was due to practicing within their established personal parameters. Providers noted their comfort in prescribing over telemedicine was attributable to their specialty and providing care to patients who require specific types of medications that providers may be generally hesitant to prescribe. Research using claims data between 2020 and 2023 shows that certain specialties (ie, psychiatry) were seen to initiate treatment for alcohol use disorder over telemedicine more than others (ie, primary care physicians) [22]. Within the same study, researchers observed a different rate of telemedicine initiations dependent on the types of medications being prescribed (eg, naltrexone at 14.6% vs topiramate at 11.8%). Our results support these findings as psychiatrists are better versed in treating mental health disorders and have tacit knowledge regarding certain medications and patient symptoms. Unsurprisingly, in this study, some providers emphasized that their comfort in prescribing decreased when faced with diagnoses they have little experience treating. From a 2023 review of pandemic-era research on telemedicine use in SUD treatment in the United States, authors noted providers would vary the number of days (supply) of prescribed medication based on their judgment of each patient interaction and context [23-26]. These findings are supported by our results as providers work within their personal parameters and determine how to proceed with prescribing based on their interaction with each patient.

Additionally, providers' personal parameters likely echo current guidelines regarding telemedicine use and limitations [27,28].

Such guidelines proffered protocols and practices based on earlier telemedicine research and stricter rules and regulations [29]. Considering that guidelines emphasize compliance with state and federal laws and regulations, we may be seeing providers' personal parameters reflecting these policies even when they feel comfortable and capable of prescribing over telemedicine [30]. Updates of such guidelines would help reframe providers' understandings of telemedicine prescribing limitations and liabilities; however, lagging legal progress limits the ability for such revisions [31,32]. With the loosening of laws and regulations regarding telemedicine prescribing, new evidence has shown that providers can care for their patients effectively while maintaining positive patient outcomes [29].

Providers also emphasized the importance of access to patient's health information, especially timely updates of exams and labs needed to prescribe certain medications such as stimulants. Certainly, the content analysis showed multiple workarounds focused on streamlining the process of accessing patient health information. Understandably, providers also noted that they feel more comfortable working with established patients, where patient health history and a personal repertoire with the patient can provide context to a telemedicine visit. Interview results suggest that an established patient—via telemedicine or in-person visit—is a central factor in how providers view their needs for timely patient health updates. This parameter is relevant for in-person and telemedicine visits. Nevertheless, with the inflated fear of liability when using telemedicine, providers show increased concern when operating outside their parameters for comfort in telemedicine prescribing [33]. Similar findings show providers may feel that additional assessment, such as physical examination, reinforces their decision-making and allays liability fears [34].

Our content analysis shows that the ultimate tipping point in whether providers prescribe via telemedicine is whether the clinical situation falls within the provider's personal parameters for comfort in prescribing. The obstacles we identified in the study, such as the e-prescription platform errors, were predominantly situations that caused frustrations in the prescribing process but were insufficient reasons for providers to discontinue prescribing medications. Of course, particular barriers, such as miscommunications with pharmacies and not receiving patient examinations or laboratory tests promptly, directly impact patient safety and outcomes. Our results support prior research showing that providers and patients faced challenges with e-prescriptions where pharmacies faced a dearth of certain medications such as buprenorphine [23,35,36]. Providers emphasized the importance of patient safety with their noted solutions: the most frequently mentioned solutions revolved around streamlining the receipt of patient health information and ensuring patient access to needed medications. Given the current, typical support infrastructure in the United States, telemedicine is not as effective for treating disorders that require laboratory examinations [37]. However, telemedicine is becoming a permanent mode of health care delivery, and successful solutions for assessment, such as diagnostic testing and information access, will be a priority for providers and other stakeholders. As a result, there is a growing support infrastructure including at-home laboratory options ranging from technicians coming to your home to home tests that can be done during a telemedicine visit [38,39]. While large health institutions such as Yale School of Medicine may have such resources available for their patients [38], small clinics or solo providers face difficulties in connecting their patients to these resources, especially in relation to an individual patient's insurance coverage [39].

Our results show that providers prescribe within the scope of their personal parameters determined by perceived limitations of telemedicine and professional expertise, provider access to relevant patient health information, and liability concerns. Once providers met their individual parameters, they expressed higher comfort in prescribing over telemedicine. With an increase in legal clarity and simplification of prescribing over telemedicine, providers may feel more capable and comfortable with the process, as guidelines would reflect the loosening of restrictions [30]. These personal parameters exist in the context of current legal and regulatory considerations; therefore, making PHE telemedicine-related policy flexibilities permanent may help expand the situations in which providers feel comfortable prescribing. Such progress may also improve patient access to mental health care services—providers may feel more comfortable expanding their patient base for telemedicine, thus alleviating access and inequity issues for mental health care treatment [40].

Limitations

This study used interviews from providers recruited through nonprobability convenience sampling via the TelehealthEngage research panel. Our findings are not generalizable to all US mental health providers. It reflects the view of interview participants, including providers from solo, small private practices; statewide health systems and clinics; and other types of mental health care settings. Additionally, as we included all prescribing mental health providers, the prescribing challenges, workarounds, and level of comfort we captured in the study may differ for other providers based on their practice focus. Future research should include a larger sample of participants including those practicing in larger health systems and academic health sciences centers, who were underrepresented in our sample.

Conclusions

Through content analysis of semistructured provider interviews, we determined that the participating providers feel comfortable prescribing over telemedicine when they practice within perceived limits of telemedicine and their own expertise, with sufficient access to patient health information, and when they do not have liability concerns. Providers mentioned multiple hindrances to prescribing over telemedicine. However, the ultimate reason they refer patients to other providers or convert encounters from telemedicine to in-person visits is that their comfort parameters are not met, and thus, patient safety is at risk. Future research, including inductive-deductive qualitative analysis of the interviews, will help create a more robust understanding of provider perspectives on telemedicine prescribing and inform future implementation and policy of prescribing over telemedicine.

Acknowledgments

We gratefully acknowledge Beju Shah for his assistance in developing the interview guide. Funding for this research was provided through Doxy.me, Inc.

Conflicts of Interest

BMW is a shareholder, and JI, HS, TO, BEB, EL, and MRC are employees of Doxy.me Inc, a commercial telemedicine company.

Multimedia Appendix 1
Semistructured interview guide.
[DOCX File , 15 KB - humanfactors_v12i1e65419_app1.docx]

References

1. Fiscal year 2024: budget in brief. U.S. Department of Health and Human Services. URL: https://www.hhs.gov/sites/default/files/fy-2024-budget-in-brief.pdf [accessed 2024-02-01]
2. CARES Act: AMA COVID-19 pandemic telehealth fact sheet. American Medical Association. URL: https://www.ama-assn.org/delivering-care/public-health/cares-act-ama-covid-19-pandemic-telehealth-fact-sheet [accessed 2024-04-05]
3. VanderWerf M, Bernard J, Barta DT, Berg J, Collins T, Dowdy M, et al. Pandemic action plan policy and regulatory summary telehealth policy and regulatory considerations during a pandemic. Telemed e-Health 2022;28(4):457-466 [FREE Full text] [doi: 10.1089/tmj.2021.0216] [Medline: 34265216]
4. DEA hosts public listening sessions on telemedicine regulations. United States Drug Enforcement Administration. URL: https://www.dea.gov/press-releases/2023/08/07/dea-hosts-public-listening-sessions-telemedicine-regulations [accessed 2023-09-21]
5. DEA extends telemedicine flexibilities for prescribing of controlled medications: second time is the charm. Foley. URL: https://www.foley.com/insights/publications/2023/10/dea-extends-telemedicine-prescribing-controlled/ [accessed 2024-04-10]
6. Huskamp HA, Riedel L, Uscher-Pines L, Busch AB, Barnett ML, Raja P, et al. Initiating opioid use disorder medication via telemedicine during COVID-19: implications for proposed reforms to the ryan haight act. J Gen Intern Med 2022;37(1):162-167 [FREE Full text] [doi: 10.1007/s11606-021-07174-w] [Medline: 34713386]
7. Niyibizi A, Haveric A, Irio G. Telehealth in opioid use disorder treatment: policy considerations for expanding access to care. J Osteopath Med 2024;124(4):179-181 [FREE Full text] [doi: 10.1515/jom-2023-0215] [Medline: 37976470]
8. Hailu R, Mehrotra A, Huskamp HA, Busch AB, Barnett ML. Telemedicine use and quality of opioid use disorder treatment in the US during the COVID-19 pandemic. JAMA Netw Open 2023;6(1):e2252381 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.52381] [Medline: 36692880]
9. Mattocks KM, Moore DT, Wischik DL, Lazar CM, Rosen MI. Understanding opportunities and challenges with telemedicine-delivered buprenorphine during the COVID-19 pandemic. J Subst Abuse Treat 2022;139:108777 [FREE Full text] [doi: 10.1016/j.jsat.2022.108777] [Medline: 35346533]
10. Lin LA, Fortney JC, Bohnert AS, Coughlin LN, Zhang L, Piette JD. Comparing telemedicine to in-person buprenorphine treatment in U.S. veterans with opioid use disorder. J Subst Abuse Treat 2022;133:108492 [FREE Full text] [doi: 10.1016/j.jsat.2021.108492] [Medline: 34175175]
11. Guillen AG, Reddy M, Saadat S, Chakravarthy B. Utilization of telehealth solutions for patients with opioid use disorder using buprenorphine: a scoping review. Telemed e-Health 2022;28(6):761-767. [doi: 10.1089/tmj.2021.0308] [Medline: 34714172]
12. Hammerslag LR, Mack A, Chandler RK, Fanucchi LC, Feaster DJ, LaRochelle MR, et al. Telemedicine buprenorphine initiation and retention in opioid use disorder treatment for medicaid enrollees. JAMA Netw Open 2023;6(10):e2336914 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.36914] [Medline: 37851446]
13. Uscher-Pines L, Parks AM, Sousa J, Raja P, Mehrotra A, Huskamp HA, et al. Appropriateness of telemedicine versus in-person care: a qualitative exploration of psychiatrists' decision making. Psychiatr Serv 2022;73(8):849-855 [FREE Full text] [doi: 10.1176/appi.ps.202100519] [Medline: 35080419]
14. Cummins MR, Ivanova J, Ong T, Soni H, Barrera JF, Wilczewski H, et al. Will the United States pass on telemedicine progress? JAMIA Open 2024;7(1):ooae016 [FREE Full text] [doi: 10.1093/jamiaopen/ooae016] [Medline: 38410742]
15. Meese KA, Hall AG, Feldman SS, Colón-López A, Rogers DA, Singh JA. Physician, nurse, and advanced practice provider perspectives on the rapid transition to inpatient and outpatient telemedicine. Telemed Rep 2022;3(1):7-14 [FREE Full text] [doi: 10.1089/tmr.2021.0034] [Medline: 35720449]
16. Telemedicine can change care for the better—with the right rules. Harvard Medical School. URL: https://hms.harvard.edu/news/telemedicine-can-change-care-better-right-rules [accessed 2024-04-25]

17. Cummins M, Ivanova J, Soni H, Robbins Z, Bunnell BE, López E, et al. Telemedicine prescribing: a national, cross-sectional survey of U.S. mental health care providers. JMIR Preprints Preprint posted online on December 21, 2024 [FREE Full text] [doi: 10.2196/preprints.63251]
18. TelehealthEngage. Doxy.me. URL: https://doxy.me/en/ [accessed 2025-01-11]
19. Bernard HR, Wutich A, Ryan GW. Analyzing Qualitative Data: Systematic Approaches. Thousand Oaks, CA: SAGE Publications; 2016.
20. Campbell JL, Quincy C, Osserman J, Pedersen OK. Coding in-depth semistructured interviews. Sociol Methods Res 2013;42(3):294-320. [doi: 10.1177/0049124113500475]
21. Raskind IG, Shelton RC, Comeau DL, Cooper HLF, Griffith DM, Kegler MC. A review of qualitative data analysis practices in health education and health behavior research. Health Educ Behav 2019;46(1):32-39 [FREE Full text] [doi: 10.1177/1090198118795019] [Medline: 30227078]
22. Huskamp HA, Uscher-Pines L, Raja P, Normand ST, Mehrotra A, Busch AB. Telemedicine for initiation of alcohol use disorder medications. JAMA Netw Open 2024;7(9):e2431594 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.31594] [Medline: 39230907]
23. Lin C, Pham H, Zhu Y, Clingan SE, Lin L, Murphy SM, et al. Telemedicine along the cascade of care for substance use disorders during the COVID-19 pandemic in the United States. Drug Alcohol Depend 2023;242:109711 [FREE Full text] [doi: 10.1016/j.drugalcdep.2022.109711] [Medline: 36462230]
24. Uscher-Pines L, Sousa J, Raja P, Mehrotra A, Barnett M, Huskamp HA. Treatment of opioid use disorder during COVID-19: experiences of clinicians transitioning to telemedicine. J Subst Abuse Treat 2020;118:108124 [FREE Full text] [doi: 10.1016/j.jsat.2020.108124] [Medline: 32893047]
25. Nordeck CD, Buresh M, Krawczyk N, Fingerhood M, Agus D. Adapting a low-threshold buprenorphine program for vulnerable populations during the COVID-19 pandemic. J Addict Med 2021;15(5):364-369 [FREE Full text] [doi: 10.1097/ADM.0000000000000774] [Medline: 33177436]
26. Duncan A, Sanders N, Schiff M, Winkelman TN. Adaptations to jail-based buprenorphine treatment during the COVID-19 pandemic. J Subst Abuse Treat 2021;121:108161 [FREE Full text] [doi: 10.1016/j.jsat.2020.108161] [Medline: 33371945]
27. Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and telepsychiatry: development of evidence-based guidance for clinicians. JMIR Ment Health 2020;7(8):e21108 [FREE Full text] [doi: 10.2196/21108] [Medline: 32658857]
28. Gough F, Budhrani S, Cohn E, Dappen A, Leenknecht C, Lewis B, et al. ATA practice guidelines for live, on-demand primary and urgent care. Telemed e-Health 2015;21(3):233-241 [FREE Full text] [doi: 10.1089/tmj.2015.0008] [Medline: 25658882]
29. Ivanova J, Ong T, Wilczewski H, Cummins M, Soni H, Barrera J, et al. Telemental health care guidelines from the first 2 years of the COVID-19 pandemic: a scoping review using content and thematic analyses. JMIR Preprints Preprint posted online on January 18, 2024 [FREE Full text] [doi: 10.2196/preprints.56534]
30. Krawczyk N, Rivera BD, King C, Dooling BCE. Pandemic telehealth flexibilities for buprenorphine treatment: a synthesis of evidence and policy implications for expanding opioid use disorder care in the United States. Health Aff Sch 2023;1(1):qxad013 [FREE Full text] [doi: 10.1093/haschl/qxad013] [Medline: 38145115]
31. Double secret protection: bridging federal and state law to protect privacy rights for telemental and mobile health users. Duke Law Journal. URL: https://dlj.law.duke.edu/article/double-secret-protection-sherman-vol67-iss5/ [accessed 2023-05-22]
32. Tong KW. Telehealth as a double-edged sword: lessons from court cases to gain understanding of medico-legal risks. Med Law 2019;38(1):85-106 [FREE Full text]
33. Nguyen M, Waller M, Pandya A, Portnoy J. A review of patient and provider satisfaction with telemedicine. Curr Allergy Asthma Rep 2020;20(11):72. [doi: 10.1007/s11882-020-00969-7] [Medline: 32959158]
34. Riley PE, Fischer JL, Nagy RE, Watson NL, McCoul ED, Tolisano AM, et al. Patient and provider satisfaction with telemedicine in otolaryngology. OTO Open 2021;5(1):2473974X20981838. [doi: 10.1177/2473974x20981838]
35. Castillo M, Conte B, Hinkes S, Mathew M, Na CJ, Norindr A, et al. Implementation of a medical student-run telemedicine program for medications for opioid use disorder during the COVID-19 pandemic. Harm Reduct J 2020;17(1):88 [FREE Full text] [doi: 10.1186/s12954-020-00438-4] [Medline: 33203460]
36. Tofighi B, McNeely J, Walzer D, Fansiwala K, Demner A, Chaudhury CS, et al. A telemedicine buprenorphine clinic to serve New York City: initial evaluation of the NYC public hospital system's initiative to expand treatment access during the COVID-19 pandemic. J Addict Med 2022;16(1):e40-e43 [FREE Full text] [doi: 10.1097/ADM.0000000000000809] [Medline: 33560696]
37. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117 [FREE Full text] [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
38. Corcoran TP. Mobile lab helps Yale pathology establish relationships, build trust with community. Yale School of Medicine. URL: https://medicine.yale.edu/pathology/news-article/mobile-lab-helps-yale-pathology-establish-relationships-build-trust-with-community/ [accessed 2024-07-10]
39. Drake K. At-home blood tests are becoming a trend. What it means for healthcare. Healthline. URL: https://www.healthline.com/health-news/at-home-blood-tests-are-becoming-a-trend-what-it-means-for-healthcare [accessed 2024-07-10]

40. Garett R, Young SD. Potential effects of digital inequality on treatment seeking for opioid use disorder. Int J Mental Health Addict 2023;21(2):890-895 [FREE Full text] [doi: 10.1007/s11469-021-00629-5] [Medline: 34466133]

Abbreviations

PHE: public health emergency
SUD: substance use disorder

Edited by A Kushniruk; submitted 14.08.24; peer-reviewed by J Ong, T Khodaveisi; comments to author 14.10.24; revised version received 10.12.24; accepted 20.12.24; published 20.03.25.

Please cite as:
Ivanova J, Cummins MR, Soni H, Ong T, Bunnell BE, López E, Welch BM
Mental Health Providers' Challenges and Solutions in Prescribing Over Telemedicine: Content Analysis of Semistructured Interviews
JMIR Hum Factors 2025;12:e65419
URL: https://humanfactors.jmir.org/2025/1/e65419
doi:10.2196/65419
PMID:

Original Paper

Health Care Professionals' Perspectives on Using eHealth Tools in Advanced Home Care: Qualitative Interview Study

Eric Vincent Rivas[1], MMSc, MD; Ulf Lesley[1], MSc, MMSc; Nadia Davoody[1], 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 18a
Stockholm, SE-171 77
Sweden
Phone: 46 852486486
Email: nadia.davoody@ki.se

Abstract

Background: The rising demand for advanced home care services, driven by an aging population and the preference for aging in place, presents both challenges and opportunities. While advanced home care can improve cost-effectiveness and patient outcomes, gaps remain in understanding how eHealth technologies can optimize these services. eHealth tools have the potential to offer personalized, coordinated care that increases patient engagement. However, research exploring health care professionals' (HCPs) perspectives on the use of eHealth tools in advanced home care and their impact on the HCP-patient relationship is limited.

Objective: This study aims to explore HCPs' perspectives on using eHealth tools in advanced home care and these tools' impact on HCP-patient relationships.

Methods: In total, 20 HCPs from 9 clinics specializing in advanced home care were interviewed using semistructured interviews. The discussions focused on their experiences with 2 eHealth tools: a mobile documentation tool and a mobile preconsultation form. The data were analyzed using content analysis to identify recurring themes.

Results: The data analysis identified one main theme: optimizing health care with eHealth; that is, enhancing care delivery and overcoming challenges for future health care. Two subthemes emerged: (1) enhancing care delivery, collaboration, and overcoming adoption barriers and (2) streamlining implementation and advancing eHealth tools for future health care delivery. Five categories were also identified: (1) positive experiences and benefits, (2) interactions between HCPs and patients, (3) challenges and difficulties with eHealth tools, (4) integration into the daily workflow, and (5) future directions. Most HCPs expressed positive experiences with the mobile documentation tool, highlighting improved efficiency, documentation quality, and patient safety. While all found the mobile preconsultation form beneficial, patient-related factors limited its utility. Regarding HCP-patient relationships, interactions with patients remained unchanged with the implementation of both tools. HCPs successfully maintained their interpersonal skills and patient-centered approach while integrating eHealth tools into their practice. The tools allowed more focused, in-depth discussions, enhancing patient engagement without affecting relationships. Difficulties with the tools originated from tool-related issues, organizational challenges, or patient-related complexities, occasionally affecting the time available for direct patient interaction.

Conclusions: The study underscores the importance of eHealth tools in enhancing advanced home care while maintaining the HCP-patient relationship. While eHealth tools modify care delivery techniques, they do not impact the core dynamics of the relationships between HCPs and patients. While most of the HCPs in the study had a positive attitude toward using the eHealth tools, understanding the challenges they encounter is crucial for improving user acceptance and success in implementation. Future development should focus on features that not only improve efficiency but also actively enhance HCP-patient relationships, such as facilitating more meaningful interactions and supporting personalized care in the advanced home care setting.

(JMIR Hum Factors 2025;12:e60582) doi:10.2196/60582

KEYWORDS

eHealth; mobile health; mHealth; advanced home care; content analysis; nurse; staff-patient relationship; aging population; patient engagement; personalized care; patient experience

Introduction

Advanced Home Care

Advanced home care, also known as advanced health care in the home, involves providing medical care or treatment directly at a patient's home, serving as an alternative care to inpatient care. The World Health Organization defines home care as "an array of health and social support services provided to clients in their residence. Such coordinated services may prevent, delay, or be a substitute for temporary or long-term institutional care" [1]. According to a study by Barakat et al [2], the components of home care include preventive actions and assessments, postdischarge actions, and evaluations, with objectives focused on enhancing or maintaining the quality of life, optimizing functional health status, and promoting independence. The composition of health care professionals (HCPs) providing home care varies depending on the patient's needs and the extent of services they require. This team may involve various HCPs, such as physicians, nurses, assistant nurses, physical therapists, occupational therapists, speech-language pathologists, home health aides, home infusion nurses (specialized nurses who administer intravenous medications in patients' homes), hospice caregivers, and medical social workers [3]. Patients receiving home care may have a wide range of health conditions and diseases, including chronic conditions, such as diabetes, hypertension, and heart disease; neurological conditions, such as Parkinson disease, dementia, and stroke; respiratory conditions, such as chronic obstructive pulmonary disease, asthma, and sleep apnea; cancer; individuals requiring wound care; and those in need of palliative care [3].

The increasing demand for home care services is driven not only by an aging population but also by the growing preference among many individuals to age in place, remaining in their own homes. The primary goal of advanced home care is to enhance patients' quality of life and clinical outcomes while simultaneously reducing health care costs and hospital readmissions [4].

Use of Electronic Health in Advanced Home Care

Organizational suppliers of health care are increasingly delivering medical care directly to patients' homes, which has created a growing reliance on medical technologies and eHealth tools [1] to support health care staff in managing not only older people care but also palliative and end-of-life care and patients with multiple comorbidities or complex diseases. The increased presence of eHealth tools also highlights the importance of considering the competencies and requirements of HCPs for using eHealth tools in older people care [2].

eHealth's vital role in advanced home care includes using telehealth, remote patient monitoring, and mobile health apps to deliver health care services to patients with complex medical needs in their homes [5]. eHealth tools facilitate communication between patients and HCPs, support medication management, monitor vital signs and symptoms, and offer educational resources to patients and caregivers. Furthermore, integrating eHealth into advanced home care can also enable organizational suppliers of health to deliver personalized and coordinated care, increase patient engagement and self-management, and improve overall patient satisfaction [2].

In Sweden, the national policy emphasizes home assistance and home care over institutionalized care, with various programs supporting long-term care for older people, assistance for people with disabilities, end-of-life care, palliative care in hospitals or hospices, and advanced palliative home care administered by municipalities. In 2017, home care met 72% of long-term care needs, while institutions served the remaining 28% [6]. Given the substantial reliance on home care for health care delivery, there is a significant need for resources, making eHealth services, such as those mentioned earlier, highly beneficial [7].

A study by Rydenfält et al [8] outlined the implementation of eHealth services in Swedish home care nursing, listing national patient summaries, mobile documentation, digital locks, digital medical lists, digital security alarms, and camera supervision as the most commonly used services. Organizational suppliers of health care had implemented mobile documentation most widely among these services at the time of the study, with plans for its continued use in the future for home care nursing.

Traditional Documentation

Health care staff have long used traditional methods, such as pen-and-paper notes, to document patient encounters, including observations, assessments, signs, symptoms, and communications during clinic consultations and home care visits. While some remnants of this traditional documentation method may persist in current practices, health care organizations have adopted information and communication technology solutions [9] and notable mobile communications [10] to address the inherent challenges of traditional documentation. Traditional documentation methods occasionally lead HCPs to omit crucial information inadvertently, record data inaccurately, produce illegible handwriting, and misunderstand patients' accounts, while requiring significant time for documentation [11].

Patient Experiences With eHealth

Researchers have extensively explored patient experiences with eHealth, particularly in home and palliative care settings, and have revealed a range of benefits and challenges. Steindal et al [3] found that telehealth applications in palliative home care enhance access to HCPs, helping patients feel more secure and safe. Similarly, Widberg et al [4] highlighted that eHealth applications facilitate better communication between patients and HCPs in palliative care. Karlsen et al [12] focused on older adults' use of telecare in home care services and emphasized their desire to age in place. These studies show that patients view technology as a valuable tool to achieve this goal, although

some resist adopting new technologies, struggle with digital proficiency, and worry about privacy and security [5].

Older patients face notable challenges when using eHealth tools. These include their preference for direct contact with HCPs, age-related cognitive decline, the stigmatization of telemonitoring devices, and the potential loss of social interactions, which are essential for well-being [5,13]. Experts emphasize the importance of balancing in-person care with eHealth use for certain populations, such as older patients with chronic conditions. Despite these obstacles, studies highlight that eHealth improves care accessibility and patient autonomy [3,5,12].

Specific eHealth applications directly benefit patients. For instance, an eHealth app tested in home care settings allowed patients to report health concerns and increased their sense of security [10] by reducing the frequency of phone calls to nurses, enabling them to prioritize care based on the app's alerts. However, frequent reporting through the app led HCPs to perceive minor health issues as significant, which increased their workload [10]. These findings illustrate the nuanced impact of eHealth on patient experiences and underline the need to carefully address patients' needs and contexts.

The Acceptance and Perspectives of HCPs on eHealth

HCPs play a pivotal role in adopting and successfully implementing eHealth solutions, as their acceptance and perspectives directly influence outcomes. Li et al [14] identified key factors that affect HCPs' acceptance of eHealth. These factors include HCP characteristics, voluntariness of use, performance and effort expectancy, and how organizational conditions either support or hinder eHealth adoption. HCPs emphasize patient autonomy, personalization, and continuity of professional support as facilitators of eHealth adoption, while they note advanced patient age as a primary barrier [15].

Qualitative studies offer additional insights into HCPs' attitudes toward eHealth. In an interview study, general practitioners expressed positive attitudes toward eHealth as a means to promote healthy lifestyles for patients and themselves [16]. They showed confidence in transitioning from traditional paper-based approaches to digital solutions, particularly those incorporating patient-reported outcome measures. Similarly, Das et al [17] examined the impact of an eHealth portal on HCP-patient interactions and reported that HCPs value the portal for its ability to provide comprehensive patient information, foster accountability, and serve as a clinical tool. However, they also identified organizational challenges, such as a lack of incentives and time constraints, which hinder eHealth integration.

Understanding HCPs' experiences and perceptions is essential for optimizing eHealth tools. While eHealth enhances health care quality, efficiency, and accessibility, its success depends on addressing the barriers that HCPs face. Targeted training and strategies to overcome organizational constraints can improve HCP satisfaction and ensure seamless integration of eHealth into clinical practice [18].

Aim of the Study

Despite extensive research in eHealth, there remains a limited understanding of how eHealth tools affect the relationship between HCPs and patients in advanced home care settings. While some studies have explored patient perspectives, fewer researchers have focused on HCPs' experiences and perceptions.

The aim of this study is to explore HCPs' perspectives on using eHealth tools in advanced home care and their impact on HCP-patient relationships. This research is particularly relevant in the context of Sweden's robust home care system.

Methods

Methodology Overview

This study used qualitative research methods, collecting data through semistructured interviews with HCPs working in advanced home care at Aleris in Sweden. Aleris is one of the major private organizational suppliers of health care in Sweden, offering a range of medical services, including advanced home care and other specialist medical treatments [19]. This approach suited the study's objective, which aimed to gain a deep understanding of HCPs' perspectives on using eHealth tools in advanced home care and their impact on relationships with patients. The flexibility of qualitative research allowed nuanced meanings to emerge from participants, captured through direct quotations from interview transcripts. The study adopted an exploratory design to further explain the role of eHealth tools in advanced home care as perceived by HCPs and examined how these tools influence their interactions with patients. This design proved beneficial in uncovering connections between ideas that were either underrepresented or not previously demonstrated in the literature [20]. An inductive approach [21] analyzed the data, aligning with the study's objectives.

The eHealth Tools

The eHealth tools used in the study are part of the SwipeCare health care process management system, which partially integrates into TakeCare, the electronic health record (EHR) system used by HCPs in their daily practice. TakeCare serves as the main repository for patient information and is used for routine clinical documentation. SwipeCare functions as an overlay on TakeCare, enhancing the existing EHR system by adding functionalities, such as iPad (Apple Inc)-based questionnaires, calculations, creating and sending questionnaires to patients, and patient reminders [22]. SwipeCare controls and coordinates the patient care process from diagnosis and treatment to aftercare and follow-up. It allows the organizational suppliers of health to define and manage the entire health care process, initiating actions within that process, some performed by SwipeCare itself (like sending out questionnaires) and others by HCPs or other units. SwipeCare interacts with TakeCare only when it needs to read from or write to the patient's record, thus complementing rather than replacing the EHR.

The 2 tools chosen for the evaluation were a mobile documentation tool and a mobile preconsultation form. The tools were chosen for their complementary nature and potential to capture a wide range of HCP-patient interactions. While the mobile documentation tool was used by the HCP, or

collaboratively by HCP and patients, during in-person visits, the mobile preconsultation form was completed independently by patients at home. This diversity in tool use allowed a comprehensive exploration of eHealth's impact on HCP-patient relationships across different contexts.

The iPad-based mobile documentation tool consists of different predefined standardized forms, which HCPs, either themselves or together with the patient, fill in. Such a form could, for example, cover the required information related to the admission of the patient to a ward or a checklist for evaluating the patient's physical status after an adjustment in the treatment. The questions in the form consist of predefined categories designed to guide HCPs in documenting patient information. Caregivers have the discretion to determine which answers are available for HCPs to fill in. In some cases, there is a designated "comments" section where HCPs can provide additional explanations or insights. However, in other instances, caregivers may choose to limit responses to predefined alternatives. This limitation mirrored traditional documentation methods, such as pen-and-paper systems, where options were similarly restricted. In addition, certain calculations, such as the Mini Nutritional Assessment (MNA) [23] or the National Early Warning Score 2 (NEWS2) [24], necessitate numerical input and cannot accommodate text responses. This further underscored the importance of predefined answer alternatives within the tool. In addition, the forms ensure no relevant aspects are missed, as questions can be made mandatory to answer.

The second tool, the mobile preconsultation form, is part of the patient portal in the system, which enables the patient to access their own portal, or web page, on their mobile or equivalent digital device. The patient first receives an SMS text message with a link to their patient portal. The patient logs in using a secure method and can then access the preconsultation form. The patient answers the questions at their own pace and time and finally sends the answers back to the caregiver. This enables HCPs to review the status of the patient before traveling to the patient's home, avoiding the time spent on questioning and filling in answers. Instead, it allows that time to be spent caring for the patient where the patient has its needs, for example, pain in the foot or problems with digestion. The intention is to free HCPs from administration to devote time to the care of the patient and encourage patients to take a more active role in their own care.

Data Collection and Participant Selection

The recruitment process involved contacting selected participants via email, facilitated by managers from each of the 9 participating clinics in Stockholm, Sweden. Clinics were selected to ensure geographic diversity across the Stockholm region, including clinics from the north, south, and other areas. Clinic heads collaborated in identifying suitable participants with experience using the eHealth tools. This selection strategy aimed to capture a comprehensive range of perspectives, strengthening the validity of the findings. Inclusion criteria for participants comprised individuals aged ≥18 years, officially employed at these clinics, and experienced in using the eHealth tools in their professional roles. The diversity in HCPs' roles and experience levels allowed for capturing a broad view of eHealth tool use in advanced home care.

Interviews took place between March and May 2023. Participants received the informed consent form via email before the interviews, which they reviewed and signed in person at the beginning of the interview session. Audio recordings and notes documented each 30- to 45-minute interview. A structured interview guide provided consistency while allowing the interviewer flexibility to explore topics relevant to HCPs' perspectives further. For example, one question from the interview guide asked participants to describe how they adapted their use of eHealth tools based on individual patient needs, providing deeper insights into their experiences and perceptions regarding the patient-HCP relationship.

The characteristics of the 20 HCPs who participated in the interviews varied across the 9 clinics. Each clinic contributed between 1 and 4 HCPs, with a median of 2, ensuring a good spread of perspectives across the clinics involved. Participants were aged between 24 to 59 years, with a mean of 45 (SD 9.3; IQR 41-52) years.

Table 1 presents participants' characteristics, categorized by profession, years of practice, duration of using mobile documentation tools, and duration of using mobile preconsultation forms. All (20/20, 100%) participants used the mobile documentation tool, while only 60% (12/20) of them used the mobile preconsultation form. This difference in tool use reflects the phased implementation approach common in health care settings, allowing for the gradual adoption and refinement of new technologies. The relatively short duration of eHealth tool use for many participants (0-2 years) illustrates the dynamic nature of health care staffing, characterized by a high turnover rate [25]. HCPs joining from other organizational suppliers of health care or units typically lacked previous experience with these specific eHealth tools. This limited experience duration accurately represents the reality found in many health care units and provides valuable insights into the initial adoption phase of eHealth tools in advanced home care.

These factors—the selection process, varied tool familiarity, and relatively short use duration—shaped the findings by offering a realistic snapshot of eHealth tool adoption in health care units where staff turnover is common. While this scenario limits observations on long-term impacts, the study offers a pragmatic view of eHealth tool implementation, helping stakeholders set realistic expectations and develop strategies that account for the realities of staff turnover and varying levels of tool familiarity in health care settings.

Table 1. Participant characteristics.

Participant	Duration of clinical practice[a] (y)	Duration of the use of mobile documentation tools (y)	Duration of the use of mobile preconsultation form (y)
Nurse 1	6-10	0-2	—[b]
Nurse 2	16-20	0-2	—
Nurse 3	6-10	0-2	0-2
Nurse 4	6-10	0-2	0-2
Nurse 5	6-10	0-2	0-2
Nurse 6	0-5	0-2	0-2
Nurse 7	6-10	3-5	3-5
Nurse 8	0-5	0-2	0-2
Nurse 9	0-5	3-5	3-5
Nurse 10	0-5	0-2	0-2
Nurse 11	6-10	>5	0-2
Nurse 12	11-15	0-2	—
Nurse 13	0-5	0-2	0-2
Nurse 14	6-10	0-2	0-2
Nurse 15	0-5	0-2	0-2
Assistant nurse 1	6-10	0-2	—
Assistant nurse 2	6-10	0-2	—
Assistant nurse 3	0-5	0-2	—
Assistant nurse 4	0-5	3-5	—
Operational therapist 1	>20	0-2	—

[a]Clinical practice: total time spent as health care professionals across all settings, including advanced home care clinics.

[b]Not applicable.

Data Analysis

The interview, recorded in audio format, underwent verbatim transcription and subsequent review for accuracy and completeness. Following the transcription of the data, participants' responses were systematically organized and categorized. The analysis was conducted using a content analysis approach grounded in the methodology described by Graneheim and Lundman [26]. The process began with a thorough data familiarization. To gain a deep understanding of its content, we immersed ourselves in the material, reading it carefully multiple times. Once a comprehensive understanding of the data was achieved, we identified *meaning units*—specific segments of text that conveyed key information relevant to the study's objectives. These meaning units were then condensed to retain their core meaning while eliminating unnecessary details. Each condensed meaning unit was assigned a code, which served as a concise label capturing its essential message. The codes were reviewed and grouped into *subcategories* based on shared patterns and similarities, providing a more detailed and refined understanding of the data. These subcategories were subsequently combined into broader *categories* that represented the distinct dimensions of our study. We then merged related categories and identified key subthemes. Ultimately, we consolidated these findings into an overarching theme. This theme captured the core insights from the data, highlighting both the benefits and challenges of eHealth tool use in advanced home care from HCPs' perspectives.

Ethical Considerations

This research was conducted in Sweden. According to the Swedish Ethical Review Act (SFS 2003:460) [27,28], this study does not require ethics approval as it does not involve the handling of sensitive personal information, as defined by the European General Data Protection Regulation (GDPR), Regulation (EU) 2016/679 [29]. However, we acknowledge that ethical principles still apply and have been followed in accordance with relevant regulations and research guidelines.

Prospective study participants received an invitation email outlining the study's purpose. Those who agreed to participate were contacted via email, SMS text messages, or phone calls to schedule interviews according to their preferences. Before beginning the interviews, participants were provided with a comprehensive explanation of the study objectives, interview procedures, and their rights. Both verbal and written consent were obtained from each participant, with informed consent signed by the participant and the researcher. The consent form included information regarding the aim of the study, potential risks and discomforts, potential benefits, how confidentiality would be handled, the right to withdraw, and consent

procedures. It also assured participants of their confidentiality. All participants were informed of their right to withdraw from the study at any stage. Furthermore, no monetary or nonmonetary compensation was provided to any participants. Finally, they were assured that all raw data, comprising audio recordings and interview notes, would be stored on a secured laptop with biometrics log-in (fingerprint) for enhanced data security.

Results

Overview of Data Analysis

The data analysis identified one main theme—*optimizing health care with eHealth: enhancing care delivery and overcoming challenges for future health care*. Two subthemes emerged: (1) enhancing care delivery, collaboration, and overcoming adoption barriers and (2) streamlining implementation and advancing eHealth tools for future health care delivery. Five categories were also identified: (1) positive experiences and benefits, (2) interactions between HCPs and patients, (3) challenges and difficulties with eHealth tools, (4) integration into the daily workflow, and (5) future directions. Table 2 shows the subcategories, categories, subthemes, and themes that emerged after coding.

The overarching theme highlighted the study's central focus on the HCPs' perspectives on the use and integration of eHealth tools in their daily practice. It emphasized how these tools impact the dynamics of health care delivery, including their influence on the relationships between HCPs and patients, and how they shape the overall patient care experience.

Table 2. Overview of the *optimizing health care with electronic health: enhancing care delivery and overcoming challenges for future health care* theme, subthemes, categories, and subcategories.

Subtheme and categories	Subcategories
Enhancing care delivery, collaboration, and overcoming adoption barriers	
Positive experiences and benefits	• Efficiency • Better documentation and enhanced patient safety
Interactions between HCPs[a] and patients	• Unaffected HCP-patient relationships • Enhanced patient engagement • Communication barrier
Challenges and difficulties with eHealth tools	• Technical and resource challenges • Patient-related factors
Streamlining implementation and advancing eHealth tools for future health care delivery	
Integration into the daily workflow	• Ease of use • System integration and application
Future directions	• Staff recommendations • Continuity of use

[a]HCP: health care professional.

Enhancing Care Delivery, Collaboration, and Overcoming Adoption Barriers

Positive Experiences and Benefits

The participants highlighted several benefits of using the mobile documentation tool. One significant advantage was the decreased documentation time during patient consultations. The mobile tool allows HCPs to document directly during patient encounters, which increase *efficiency* by eliminating the need for double documentation [30] and reducing the time spent on paperwork. Approximately two-thirds (13/20, 65%) of the HCPs indicated that the mobile documentation tool helped them save time on documentation:

> *I think it [the mobile documentation tool] decreases my work time [taking clinical notes during consultation].* [Nurse 6]

> *It saves me time because I document while I am at the patient's house.* [Nurse 9]

Another benefit mentioned was the standardization of content for clinical note-taking. The use of standardized templates within the mobile tool ensured that all relevant information was captured consistently, which is crucial for effective patient care. Participants also expressed an enhanced sense of safety and security for patients when using the mobile documentation tool. This improvement was attributed to more accurate and timely documentation of health status, which contributed to better patient outcomes. The time saved through efficient documentation potentially translated to improved patient care quality. With reduced administrative burden, HCPs could devote more time to meaningful patient interactions, allowing more in-depth discussions about health concerns and treatment plans.

While most (13/20, 65%) participants perceived the mobile documentation tool as beneficial, some noted challenges. A few (7/20, 35%) participants reported that it disrupted their routine workflow by prolonging consultation times. They found that carrying the iPad during home care visits added to their equipment load, making it inconvenient:

> *I think it [the mobile documentation tool] increases my work time. My usual consultation is about 20-30 minutes, with it [the mobile documentation tool], it can take 45-50 minutes. We carry a lot of equipment during home care visits, including a laptop, and it can be heavy. The iPad is an additional weight.* [Nurse 1]

> *Because we cannot do everything we need in the [mobile documentation] tool, it kind of doubles the work for us.* [Nurse 4]

Among users of the mobile preconsultation form, participants highlighted several advantages, including knowing the patient's medical concerns in advance and saving time on preparations. This previsit work done by patients reduced the administrative workload on HCPs, allowing for a more efficient use of consultation time.

Standardizing documentation through the mobile documentation tool and preconsultation form may lead to more consistent and thorough patient assessments across different HCPs. This standardization ensures that all relevant information is captured systematically, potentially improving the continuity of care and reducing the risk of overlooking critical patient data.

The use of eHealth tools in advanced home care led to *better documentation practices while enhancing patient safety*. The majority (13/20, 65%) of the participants found the standardized content of the mobile documentation tool beneficial, noting that it ensured they would not overlook critical information:

> *It is kind of nice that the [mobile documentation] tool has a lot of information that can help with taking notes. Also, the setup is the same for everyone, so I think the note-taking is sort of standardized.* [Nurse 1]

This standardized approach to documentation not only improved consistency across different HCPs but also potentially enhanced the quality of patient care. By ensuring comprehensive data collection, the tool might lead to more informed clinical decision-making and improved continuity of care. The structured format guided HCPs through specific assessment points, helping to ensure that all relevant aspects of patient care were addressed during consultations. Improved documentation through eHealth tools contributed to patient safety and security. Participants noted that accurate information recorded in EHRs reassures patients, helping them feel more secure in their care:

> *When we use [the mobile documentation tool], we can check the necessary things we have to ask the patient. We will not forget anything because we have a guide, I think that is good for the patient...it can increase patient security.* [Nurse 13]

> *I think they [patients] feel safer when I can write everything in the [mobile documentation] tool while I am with them, and they can check what I write.* [Nurse 8]

The ability for patients to review clinical notes during consultations further enhanced their confidence, fostering a sense of transparency in the health care process and further strengthening the relationship. This aligned with the principle of "equal care" in Sweden, which emphasizes the importance of providing consistent quality care to all patients, regardless of their location or the HCPs they encounter [31].

By standardizing documentation practices, eHealth tools contributed to this principle by ensuring that all relevant patient information is consistently captured and accessible to different HCPs. This standardization not only improved continuity of care but also enhanced trust between patients and HCPs. The structured format of the mobile documentation tool facilitated comprehensive data collection, allowing HCPs to focus on critical issues during consultations. This approach optimized patient interactions and strengthened the overall quality of care provided, aligning with Sweden's commitment to equitable health care access.

Interactions Between HCPs and Patients

All (20/20, 100%) participants indicated that their *relationship with patients remained unaffected* by the implementation of both the mobile documentation tool and the mobile preconsultation form. This finding suggested that HCPs were able to maintain their interpersonal skills and patient-centered approach while integrating eHealth tools into their practice:

> *I think nothing has changed in how I interact with patients when using it [the mobile documentation tool].* [Nurse 5]

The consistency in relationships despite introducing new technology highlighted the adaptability of HCPs in maintaining effective communication with patients. While the eHealth tools introduced new elements to patient interactions, such as the physical presence of devices, HCPs appeared to successfully navigate these changes without compromising the quality of their patient relationships. Furthermore, the mobile preconsultation form may have enhanced patient engagement by allowing patients to provide more comprehensive information before consultations. This previsit work could potentially lead to more focused and in-depth discussions during face-to-face interactions, without negatively impacting the HCP-patient relationship.

The mobile preconsultation form emerged as a tool that potentially *enhances patient engagement* in their care. Some (3/12, 25%) participants noted that certain patients provided more comprehensive information through written responses on the mobile preconsultation form than during face-to-face interactions with the HCPs. This suggested that the tool may lead to more informed and collaborative interactions during consultations:

> *If you send out the form, I think you can get more information from the patients because they [patients] have more time to think about it rather than when we ask them in person.* [Nurse 7]

The previsit work done by patients allowed them to elaborate on both current problems and positive aspects of their health status. This more comprehensive written information might shift the dynamics of face-to-face interactions, potentially allowing more focused and in-depth discussions during consultations. Furthermore, the mobile preconsultation form's ability to highlight patients' medical concerns in advance

enabled HCPs to prepare more effectively for consultations. This preparation may contribute to more efficient and patient-centered interactions, as evidenced by one participant's comment:

> *We send out the form once every four weeks for when we do the team rounds, it can help us in doing the team rounds, making it faster and more efficient, in my experience.* [Nurse 5]

By facilitating more comprehensive information gathering and allowing HCPs to focus on critical issues, the mobile preconsultation form may enhance patient engagement and potentially improve the quality of HCP-patient interactions in advanced home care settings.

Participants mentioned *communication barrier* as a challenge. While the eHealth tools offered numerous benefits, 20% (4/20) of the participants observed noted challenges in their use during patient interactions. These HCPs noted that the iPad (mobile documentation tool) could sometimes be a barrier to effective communication:

> *When I am holding the iPad in front of the patient, it kind of becomes a barrier between us.* [Nurse 6]

> *When using the iPad, you tend to look down instead of looking at the patient directly.* [Nurse 9]

> *Maybe some patients find it rude [using an iPad during consultation], but most of them understand that we do it to get everything right. Also, I think sometimes I have a less open conversation with patients.* [Nurse 10]

Most (16/20, 80%) participants did not report communication issues, suggesting that the majority of HCPs were able to integrate the eHealth tools without significant impact on patient interactions. Compared to traditional pen-and-paper methods, the eHealth tool presented both advantages and challenges for patient-HCP communication. While pen-and-paper notes may feel less intrusive during patient interactions, the eHealth tool offers improved legibility, standardization, and ease of data retrieval, potentially enhancing the quality of information shared with patients. However, the physical presence of the device could create a perceived barrier for some HCPs and patients. The eHealth tool's structured format ensures comprehensive documentation, potentially reducing the risk of omitting crucial information compared to freeform paper notes. This standardization may lead to more consistent and thorough patient assessments across different HCPs, indirectly benefiting patient care and communication. However, this structure might sometimes limit the flexibility in capturing nuanced patient narratives.

Regarding focus on issues and medical decision-making, the eHealth tool's ability to provide instant access to patient history and standardized assessments can potentially enhance HCP's ability to focus on critical issues during patient interactions. However, as noted by some participants, the tool might occasionally divert attention from direct patient interaction. These observations highlighted the need for balanced use of eHealth tools, combining their benefits with maintaining effective patient communication. HCPs may need to develop strategies to integrate these tools seamlessly into their patient interactions, ensuring that technology enhances rather than hinders the HCP-patient relationship.

Challenges and Difficulties With the eHealth Tools

While most (14/20, 70%) participants reported no technical issues, some HCPs encountered *technical and resource challenges* when using eHealth tools in advanced home care. Specifically, 30% (6/20) of the participants faced technical problems with the mobile documentation tool, including log-in difficulties, frequent software updates, and issues importing clinical notes to the patient's EHR:

> *First, when you come in to work, you have to log in to the [mobile documentation tool]. Sometimes, it's a problem. Now, with the frequent updates—they do updates like every week—there is often something wrong after each update...* [Nurse 3]

Notably, 100% (12/12) of the participants who used the mobile preconsultation form reported no technical issues at all. Resource limitations were a more widespread concern. All (20/20, 100%) participants noted the need to share iPads among staff members, potentially limiting access and flexibility in documentation:

> *The staff has to share iPads [with the mobile documentation tool] within the team.* [Nurse 9]

These challenges highlighted the importance of robust IT infrastructure and adequate resource allocation in eHealth implementation. While technical issues affected only a minority (6/20, 30%) of users, device sharing impacted all (20/20, 100%) participants, potentially influencing HCP efficiency and, consequently, patient care quality.

Patient-related factors, such as patient characteristics, influenced the adoption and use of eHealth tools in advanced home care. Approximately 55% (11/20) of the participants noted that the patient's age and health conditions affected receptiveness to eHealth tools, particularly the mobile preconsultation form:

> *Not all patients are open to using the [mobile preconsultation] form. Usually, older patients or those with difficult diseases are not open to it.* [Nurse 12]

HCPs demonstrated adaptability by adjusting their approach based on patient characteristics. This flexibility was crucial for maintaining quality care across diverse patient populations. However, it also raised concerns about potential disparities in care delivery between technologically proficient patients and those less comfortable with eHealth tools. Some patients expressed skepticism about using the mobile preconsultation form, particularly due to privacy concerns:

> *There are patients who are skeptical about using it [the mobile preconsultation form], as it involves using their [personal information] BankID [a Swedish electronic identification system].* [Nurse 11]

On the other hand, a subset of patients, predominantly younger individuals, demonstrated an openness to using eHealth tools, according to 55% (11/20) of the participants. This varying receptiveness highlighted the need for a personalized approach in implementing eHealth tools in advanced home care settings,

ensuring that all patients receive appropriate care regardless of their technological proficiency.

Streamlining Implementation and Advancing eHealth Tools for Future Health Care Delivery

Integration Into the Daily Workflow

Regarding *ease of use*, 75% (15/20) of the participants found the mobile documentation tool easy to use. The initial introduction and demonstration of these eHealth tools to HCPs lasted an hour and was perceived as sufficient by all the 12 (12/12, 100%) participants who received it. Some (4/20, 20%) participants had the opportunity to familiarize themselves with the tool beforehand as part of the implementation team, while 12 (12/20, 60%) participants received information during introductory seminars at their clinics. Some (4/20, 20%) participants only underwent introductory training after using the mobile documentation tool:

> *I have tried it [the mobile documentation tool], honestly, without any training... I tried to use it... I think it is easy.* [Assistant Nurse 2]

> *My colleague taught me how to use it [the mobile documentation tool].* [Nurse 8]

In addition, 25% (5/20) of the participants reported that the log-in process and navigating through all the components of the mobile documentation tool required some time to become familiar with, but they gradually improved with more use:

> *As I continue to use it [the mobile documentation tool], I learn more about it, which makes me better at using it.* [Nurse 2]

Regarding *system integration and adaption* all (20/20, 100%) participants highlighted challenges in integrating eHealth tools into existing systems and their practical *application* in advanced home care. A key limitation was the incomplete integration of the mobile documentation tool with the existing EHR system, TakeCare. While clinical notes could be uploaded to the patient's journal, other crucial data remained inaccessible through the mobile tool, requiring staff to rely on computers for a complete view of patient information:

> *It [the mobile documentation tool] doesn't have all the functions that you have in the EHR.* [Nurse 7]

> *[B]ecause with the computer, I can do and see everything on the patient's EHR, but not with the iPad.* [Nurse 4]

The lack of full system integration disrupted seamless workflows and raised concerns about potential inconsistencies in documentation and delays in accessing critical information during consultations. Participants expressed a strong desire for improved interoperability to ensure that all necessary data are accessible on the mobile device.

Regarding practical application, HCPs generally found the mobile preconsultation form useful for integrating into their existing workflows. However, its applicability was perceived as limited for certain patient groups and care scenarios. For example, HCPs noted that sending the form before every consultation was unnecessary for patients with chronic conditions requiring long-term care as their primary clinical concerns often remained consistent:

> *For example, I have had this patient for a very long time, and the patient has a chronic condition. I don't need to send the form [before] every visit because the status is still the same.* [Nurse 11]

Despite these limitations, participants acknowledged the value of the preconsultation form for new patients or those with changing health conditions. It allowed HCPs to gather detailed information in advance, facilitating more focused consultations. These findings underscored the importance of improving system integration to enhance workflow efficiency while adapting eHealth tools to meet diverse patient needs in advanced home care settings.

Future Directions

HCPs shared *recommendations* on enhancing the effectiveness of the eHealth tools in their daily workflow in advanced home care. Most (16/20, 80%) participants expressed that completely integrating the mobile documentation tool with the patient's EHR system would improve efficiency. They suggested that accessing the same information as in the patient's EHR directly on the mobile device would eliminate the need to switch between the iPad and the computer for clinical documentation, improving their workflow:

> *Lab results and checking off the to-do list, that is what is still missing, I think...and an easier login.* [Nurse 9]

All (12/12, 100%) HCPs who used the mobile preconsultation form found it adequate. However, 17% (2/12) of the participants suggested further enhancements to the interface, including languages other than Swedish:

> *The [mobile preconsultation] form does not look very good, in my opinion. It can be better.* [Nurse 11]

Regarding equipment use, half (10/20, 50%) of the participants expressed that having sufficient iPads for all staff members would be advantageous, reducing the need for frequent logging in and out.

Approximately two-thirds (14/20, 70%) of participants anticipated *continuity of use* of the 2 eHealth tools in advanced home care, acknowledging the emergence of the new technologies. In addition, 45% (9/20) of the HCPs expressed optimism regarding the increased patient acceptance of technology in health care:

> *I think they [patients] will use them [eHealth tools] more and more in the future.* [Nurse 7]

Discussion

Principal Findings

The aim of this study was to explore HCPs' perspectives on using eHealth tools in advanced home care and their impact on HCP-patient relationships. In total, 20 HCPs from 9 different advanced home care clinics participated in the interviews, leading to the identification of five categories: (1) positive experiences and benefits, (2) interactions between HCPs and

patients, (3) challenges and difficulties with the eHealth tool, (4) integration into the daily workflow, and (5) future directions.

Overall, most (13/20, 65%) HCPs expressed a positive perspective regarding the 2 eHealth tools used in the study: the mobile documentation tool and the mobile preconsultation form. The results showed that HCPs identified several benefits for themselves and their patients. Notably, 65% (13/20) of the participants reported that the mobile documentation tool allowed them to document directly during patient encounters, eliminating double documentation and saving time. Standardized templates within the tool ensured consistency in clinical note-taking, which participants believed enhanced patient safety by reducing errors and omissions.

While most (13/20, 65%) HCPs reported increased efficiency, 35% (7/20) experienced the opposite, suggesting that user acceptance of eHealth is multifaceted. These participants reported that the mobile documentation tool disrupted their routine workflow by prolonging consultation times and adding to their equipment load during home care visits. This highlights the importance of considering individual user experiences in eHealth implementation. The ease of use of the eHealth tools was a significant factor in their integration into daily workflows. Most (15/20, 75%) of the participants found the mobile documentation tool easy to use, indicating successful integration. This high level of usability likely contributed to the overall positive reception of the tools and suggested that well-designed eHealth solutions can be effectively adopted in advanced home care settings.

However, HCPs noted challenges, both tool-related and patient-related. Some (6/20, 30%) of the participants faced technical problems with the mobile documentation tool, including log-in difficulties and issues importing clinical notes to the EHR. These technical challenges highlight the need for robust IT infrastructure and ongoing technical support in eHealth implementation. Resource limitations were a widespread concern, with all participants noting the need to share iPads among staff members. This sharing potentially limits access and flexibility in documentation, impacting workflow efficiency. Future eHealth implementations should consider resource allocation to ensure adequate device availability for all HCPs.

Patient characteristics influenced the adoption of eHealth tools, particularly the mobile preconsultation form. Approximately 55% (11/20) of the participants noted that patient age and health conditions affected receptiveness to these tools. This finding underscores the importance of tailoring eHealth implementations to diverse patient populations and considering alternative approaches for patients less comfortable with technology. The mobile preconsultation form emerged as a tool that potentially enhances patient engagement. Some (3/12, 25%) of the participants using this tool noted that certain patients provide more comprehensive information through written responses than face-to-face interactions with HCPs. This previsit work allows for more focused and efficient consultations, potentially improving the quality of HCP-patient interactions.

Regarding the HCP-patient relationship, all (20/20, 100%) participants reported that the use of eHealth tools had no significant impact on the relationship with patients. This suggests that HCPs successfully maintained their interpersonal skills and patient-centered approach despite introducing new technology. However, 20% (4/20) of the participants noted that using devices such as iPads could occasionally act as a communication barrier during consultations, highlighting the need for strategies to seamlessly integrate technology into patient interactions.

Looking to the future, 70% (14/20) of the participants anticipated continued use of the eHealth tools in advanced home care, reflecting their perceived long-term value despite existing challenges. In addition, 45% (9/20) of the HCPs expressed optimism about increased patient acceptance of technology in health care. While participants did not report changes in HCP-patient interactions, they emphasized that current eHealth tools successfully maintain existing relationships without negatively impacting communication or rapport.

Comparison With Prior Work

Mathijssen et al [32] highlighted a similar finding of positive perception of eHealth by HCPs in home care. Despite recognized challenges such as inadequate training and support, data privacy concerns, and the necessity for more user-friendly technology, eHealth in home care is seen as a promise to enhance care quality. A study focusing on a specific eHealth tool, electronic medication administration records, emphasized the importance of cautious implementation due to potential unintended consequences [33]. Participants noted that the electronic medication administration records introduced new tasks alongside existing nursing tasks. In contrast, our study found no perceived changes in the workflow or work situation due to eHealth tools, potentially contributing to their overall positive reception.

Carlqvist et al [34] conducted a qualitative interview study examining how an eHealth application can serve as a value-creating resource from the perspective of HCPs. Their findings indicated that while such applications can enhance proactive communication and support patient engagement in self-care, challenges remain. Notably, patients' difficulties in using the application or performing measurements sometimes led to value destruction, requiring time-consuming recovery efforts by professionals. This contrasts with our study, where HCPs reported no significant workflow disruptions due to eHealth tools, potentially contributing to their positive acceptance.

Nakrem et al [35] explored the introduction of digital medicine dispensers in home health care services and their influence on patient-caregiver relationships. They found that while such technologies can improve efficiency and enhance patient independence, they may also strain patient-caregiver relationships if not aligned with patients' needs and safety concerns. However, in our study, HCPs did not perceive eHealth tools as affecting their relationships with patients, suggesting that the specific type of technology and its implementation context play crucial roles in acceptance and impact.

Furthermore, Ramachandran et al [36] conducted a review examining the impact of eHealth on patient-HCP and HCP-HCP relationships in primary care. They found that eHealth can have

both positive and negative effects on relationships and trust, influenced by factors, such as technology design, patient demographics, and organizational implementation strategies. This aligns with our findings, where the successful integration of eHealth tools without perceived workflow disruptions may be attributed to effective implementation strategies and user-centered design.

Future Research

Future studies could explore strategies for seamlessly integrating eHealth tools into patient interactions without compromising communication quality. In addition, research into personalized eHealth solutions that accommodate diverse patient needs and preferences could address challenges related to patient characteristics. Investigating the long-term impacts of eHealth tools on care quality, patient outcomes, HCP-patient relationships, and HCP job satisfaction would provide valuable insights for future implementations.

Another important area for investigation is the development and implementation of features that actively enhance HCP-patient relationships in advanced home care environments. For instance, improving patient education modules or creating shared decision-making tools tailored to home care settings could address some of the gaps identified in this study, such as maintaining meaningful interactions while using eHealth tools.

Furthermore, exploring the role of relatives in supporting patients with eHealth tools could offer new avenues for enhancing patient engagement and care delivery [37]. Relatives could assist with completing preconsultation forms or managing digital tools for patients who face challenges due to age or health conditions. Recognizing relatives as key stakeholders alongside HCPs may improve overall patient support in advanced home care settings [38].

Limitations

This study focused specifically on advanced home care settings in Sweden, which may limit the applicability of findings to other health care contexts. The unique characteristics of advanced home care, such as the need for mobile documentation and remote patient monitoring, shaped the experiences reported by HCPs. In addition, the exclusive focus on HCPs' interviews represents another limitation. Incorporating patients' perspectives could offer valuable insights and contribute to a more comprehensive understanding of the implementation of eHealth tools in advanced home care.

Conclusions

HCPs in advanced home care recognize the value of eHealth tools in their daily work while maintaining the quality of their relationships with patients. The study provides insights into both the experiences of HCPs using eHealth tools and how these tools influence their interactions with patients in advanced home care settings. Positive experiences related to eHealth use and the level of integration of eHealth tools into daily work motivate staff to use these tools, potentially allowing for more time and focus on patient interactions. By contrast, negative experiences with the eHealth tools limit their use and acceptance.

Notably, HCPs reported no perceived changes in their relationships with patients when using the mobile documentation tool and the preconsultation form tool. This suggests that eHealth tools can be integrated into care delivery without compromising the personal aspect of HCP-patient interactions. Although most of the staff in the study had a positive view of their use of the eHealth tools, understanding the challenges they encounter is essential to increase acceptance and success in implementation further. Future development should focus on features that not only improve efficiency but also actively enhance HCP-patient relationships in the advanced home care setting.

Acknowledgments

The authors would like to thank all the health care staff members who participated in this study for their time and valuable insights. Artificial intelligence was used for writing assistance. ND was funded by the Health Informatics Center at Karolinska Institutet. The funding did not influence the study's design, data collection, analysis, or manuscript writing.

Authors' Contributions

EVR, ND, and UL participated in the study design. Data collection was conducted by EVR. EVR performed the initial analysis, which was then reviewed and refined through discussions with ND. The first draft of the manuscript was prepared by EVR, with all authors contributing to subsequent writing and review processes.

Conflicts of Interest

During this study, UL was employed by the medical technology company providing the eHealth tool. Therefore, UL did not participate in the data collection, analysis, and manuscript writing of the study and was limited to the initial part of the design phase and preparing and reviewing the manuscript.

References

1. Ten Haken I, Ben Allouch S, van Harten WH. The use of advanced medical technologies at home: a systematic review of the literature. BMC Public Health 2018 Feb 26;18(1):284 [FREE Full text] [doi: 10.1186/s12889-018-5123-4] [Medline: 29482550]

2. Barakat A, Woolrych RD, Sixsmith A, Kearns WD, Kort HS. eHealth technology competencies for health professionals working in home care to support older adults to age in place: outcomes of a two-day collaborative workshop. Med 2 0 2013;2(2):e10 [FREE Full text] [doi: 10.2196/med20.2711] [Medline: 25075233]
3. Steindal SA, Nes AA, Godskesen TE, Dihle A, Lind S, Winger A, et al. Patients' experiences of telehealth in palliative home care: scoping review. J Med Internet Res 2020 May 05;22(5):e16218 [FREE Full text] [doi: 10.2196/16218] [Medline: 32369037]
4. Widberg C, Wiklund B, Klarare A. Patients' experiences of eHealth in palliative care: an integrative review. BMC Palliat Care 2020 Oct 14;19(1):158 [FREE Full text] [doi: 10.1186/s12904-020-00667-1] [Medline: 33054746]
5. Magdalena M, Bujnowska-Fedak M, Grata-Borkowska U. Use of telemedicine-based care for the aging and elderly: promises and pitfalls. Smart Homecare Technol Telehealth 2015 May;3:91-105 [FREE Full text] [doi: 10.2147/shtt.s59498]
6. Glenngård AH. The Swedish health care system. In: Tikkanen R, Osborn R, Mossialos E, Djordjevic A, Wharton G, editors. International Profiles of Health Care Systems. Adelaide, Australia: The Commonwealth Fund; 2020:181-189 URL: https://www.commonwealthfund.org/sites/default/files/2020-12/International_Profiles_of_Health_Care_Systems_Dec2020.pdf
7. Lindberg B, Nilsson C, Zotterman D, Söderberg S, Skär L. Using information and communication technology in home care for communication between patients, family members, and healthcare professionals: a systematic review. Int J Telemed Appl 2013;2013:461829 [FREE Full text] [doi: 10.1155/2013/461829] [Medline: 23690763]
8. Rydenfält C, Persson J, Erlingsdottir G, Johansson G. eHealth services in the near and distant future in Swedish home care nursing. Comput Inform Nurs 2019 Jul;37(7):366-372. [doi: 10.1097/CIN.0000000000000536] [Medline: 31135467]
9. Lucas B, Schladitz P, Schirrmeister W, Pliske G, Walcher F, Kulla M, et al. The way from pen and paper to electronic documentation in a German emergency department. BMC Health Serv Res 2019 Aug 09;19(1):558 [FREE Full text] [doi: 10.1186/s12913-019-4400-y] [Medline: 31399096]
10. Göransson C, Eriksson I, Ziegert K, Wengström Y, Langius-Eklöf A, Brovall M, et al. Testing an app for reporting health concerns- experiences from older people and home care nurses. Int J Older People Nurs 2018 Jun 05;13(2):e12181. [doi: 10.1111/opn.12181] [Medline: 29210218]
11. Akhu-Zaheya L, Al-Maaitah R, Bany Hani S. Quality of nursing documentation: paper-based health records versus electronic-based health records. J Clin Nurs 2018 Feb 20;27(3-4):e578-e589. [doi: 10.1111/jocn.14097] [Medline: 28981172]
12. Karlsen C, Ludvigsen MS, Moe CE, Haraldstad K, Thygesen E. Experiences of community-dwelling older adults with the use of telecare in home care services: a qualitative systematic review. JBI Database System Rev Implement Rep 2017 Dec;15(12):2913-2980. [doi: 10.11124/JBISRIR-2017-003345] [Medline: 29219874]
13. Roberts A, Philip L, Currie M, Mort A. Striking a balance between in-person care and the use of eHealth to support the older rural population with chronic pain. Int J Qual Stud Health Well-being 2015 Sep 02;10(1):27536 [FREE Full text] [doi: 10.3402/qhw.v10.27536] [Medline: 26336996]
14. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013 Apr 16;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
15. Cohen Rodrigues TR, de Buisonjé DR, Keesman M, Reijnders T, van der Geer JE, Janssen V, et al. Facilitators of and barriers to lifestyle support and eHealth solutions: interview study among health care professionals working in cardiac care. J Med Internet Res 2021 Oct 15;23(10):e25646 [FREE Full text] [doi: 10.2196/25646] [Medline: 34652280]
16. Brandt CJ, Søgaard GI, Clemensen J, Sndergaard J, Nielsen JB. General practitioners' perspective on eHealth and lifestyle change: qualitative interview study. JMIR Mhealth Uhealth 2018 Apr 17;6(4):e88 [FREE Full text] [doi: 10.2196/mhealth.8988] [Medline: 29666045]
17. Das A, Faxvaag A, Svanæs D. The impact of an eHealth portal on health care professionals' interaction with patients: qualitative study. J Med Internet Res 2015 Nov 24;17(11):e267 [FREE Full text] [doi: 10.2196/jmir.4950] [Medline: 26601678]
18. Enam A, Torres-Bonilla J, Eriksson H. Evidence-based evaluation of eHealth interventions: systematic literature review. J Med Internet Res 2018 Nov 23;20(11):e10971 [FREE Full text] [doi: 10.2196/10971] [Medline: 30470678]
19. Om Aleris. Aleris. URL: https://www.aleris.se/om-aleris/ [accessed 2024-12-16]
20. Green J, Thorogood N. Qualitative Methods for Health Research. 4th edition. Thousand Oaks, CA: Sage Publications; 2018.
21. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006 Jun 01;27(2):237-246. [doi: 10.1177/1098214005283748]
22. SwipeCare - an award-winning system for digital efficiency and automation of care. Strikersoft. URL: https://strikersoft.com/en/focusareas/ehealth/ [accessed 2025-03-10]
23. Vellas B, Guigoz Y, Garry PJ, Nourhashemi F, Bennahum D, Lauque S, et al. The Mini Nutritional Assessment (MNA) and its use in grading the nutritional state of elderly patients. Nutrition 1999 Feb;15(2):116-122. [doi: 10.1016/s0899-9007(98)00171-3] [Medline: 9990575]
24. Smith GB, Redfern OC, Pimentel MA, Gerry S, Collins GS, Malycha J, et al. The National Early Warning Score 2 (NEWS2). Clin Med (Lond) 2019 May;19(3):260 [FREE Full text] [doi: 10.7861/clinmedicine.19-3-260] [Medline: 31092526]
25. Eriksson A, Jutengren G, Dellve L. Job demands and functional resources moderating assistant and registered nurses' intention to leave. Nurs Open 2021 Mar 20;8(2):870-881 [FREE Full text] [doi: 10.1002/nop2.694] [Medline: 33570298]

26. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
27. Swedish Ethical Review Act (SFS 2003:460). Swedish Code of Statutes. 2004. URL: https://www.riksdagen.se/sv/dokument-och-lagar/dokument/svensk-forfattningssamling/lag-2003460-om-etikprovning-av-forskning-som_sfs-2003-460/ [accessed 2025-03-06]
28. What the Act says. Swedish Ethical Review Authority. 2025. URL: https://etikprovningsmyndigheten.se/en/what-the-act-says/#fl-main-content [accessed 2025-03-06]
29. EUR-Lex. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [accessed 2025-03-06]
30. Laitinen H, Kaunonen M, Åstedt-Kurki P. The impact of using electronic patient records on practices of reading and writing. Health Informatics J 2014 Dec;20(4):235-249 [FREE Full text] [doi: 10.1177/1460458213492445] [Medline: 25411220]
31. Towards a good and equitable health. Public Health Agency of Sweden. 2021. URL: https://www.folkhalsomyndigheten.se/contentassets/bb50d995b033431f9574d61992280e61/towards-good-equitable-health.pdf [accessed 2024-04-29]
32. Mathijssen E, de Lange W, Bleijenberg N, van Houwelingen T, Jaarsma T, Trappenburg J, et al. Factors that influence the use of eHealth in home care: scoping review and cross-sectional survey. J Med Internet Res 2023 Mar 09;25:e41768 [FREE Full text] [doi: 10.2196/41768] [Medline: 36892935]
33. Karnehed S, Erlandsson LK, Norell Pejner M. Nurses' perspectives on an electronic medication administration record in home health care: qualitative interview study. JMIR Nurs 2022 Apr 22;5(1):e35363 [FREE Full text] [doi: 10.2196/35363] [Medline: 35452400]
34. Carlqvist C, Hagerman H, Fellesson M, Ekstedt M, Hellström A. Health care professionals' experiences of how an eHealth application can function as a value-creating resource - a qualitative interview study. BMC Health Serv Res 2021 Nov 05;21(1):1203 [FREE Full text] [doi: 10.1186/s12913-021-07232-3] [Medline: 34740340]
35. Nakrem S, Solbjør M, Pettersen IN, Kleiven HH. Care relationships at stake? Home healthcare professionals' experiences with digital medicine dispensers - a qualitative study. BMC Health Serv Res 2018 Jan 15;18(1):26 [FREE Full text] [doi: 10.1186/s12913-018-2835-1] [Medline: 29334953]
36. Ramachandran M, Brinton C, Wiljer D, Upshur R, Gray CS. The impact of eHealth on relationships and trust in primary care: a review of reviews. BMC Prim Care 2023 Nov 03;24(1):228 [FREE Full text] [doi: 10.1186/s12875-023-02176-5] [Medline: 37919688]
37. Nilsson Å, Skär L, Söderberg S. Nurses' views of shortcomings in patent care encounters in one hospital in Sweden. J Clin Nurs 2015 Oct 16;24(19-20):2807-2814. [doi: 10.1111/jocn.12886] [Medline: 26177676]
38. Jazieh AR, Volker S, Taher S. Involving the family in patient care: a culturally tailored communication model. Glob J Qual Saf Healthc 2020;1(2):33-37. [doi: 10.4103/JQSH.JQSH_3_18]

Abbreviations

EHR: electronic health record
HCP: health care professional
MNA: Mini Nutritional Assessment
NEWS2: National Early Warning Score 2

Edited by C Jacob; submitted 17.05.24; peer-reviewed by R Marshall, C Göransson; comments to author 06.12.24; revised version received 11.01.25; accepted 07.02.25; published 24.03.25.

Please cite as:
Rivas EV, Lesley U, Davoody N
Health Care Professionals' Perspectives on Using eHealth Tools in Advanced Home Care: Qualitative Interview Study
JMIR Hum Factors 2025;12:e60582
URL: https://humanfactors.jmir.org/2025/1/e60582
doi:10.2196/60582
PMID:

Using Social Media Platforms to Raise Health Awareness and Increase Health Education in Pakistan: Structural Equation Modeling Analysis and Questionnaire Study

Malik Mamoon Munir, PhD; Nabil Ahmed, MPH

Integral Global Consulting LLC, 3541 Habersham at Northlake Parkway Road, Building E, Tucker, GA, United States

Corresponding Author:
Malik Mamoon Munir, PhD
Integral Global Consulting LLC, 3541 Habersham at Northlake Parkway Road, Building E, Tucker, GA, United States

Abstract

Background: Current health care education methods in Pakistan use traditional media (eg, television and radio), community health workers, and printed materials, which often fall short of reach and engagement among most of the population. The health care sector in Pakistan has not yet used social media effectively to raise awareness and provide education about diseases. Research on the impact social media can have on health care education in Pakistan may expand current efforts, engage a wider audience, and reduce the disease burden on health care facilities.

Objective: This study aims to evaluate the perceptions of health care professionals and paramedic staff regarding social media use to raise awareness and educate people about diseases as a potential means of reducing the disease burden in Pakistan.

Methods: The study used two-stage structural equation modeling (SEM). Data analysis used AMOS 26.0 software, adopting scales from previous literature. Four-item scales for each social media usefulness and health awareness construct and 8-item scales for health care education constructs were adopted on the basis of their higher loading in alignment with psychometric literature. A 7-point Likert scale was used to measure each item. Data collection used convenience sampling, with questionnaires distributed to more than 450 health care professionals and paramedic staff from 2 private hospitals in Lahore, Pakistan. There were 389 useful responses received. However, 340 completed questionnaires were included in the data analysis.

Results: The study found that all the squared multiple correlation (SMC) values were greater than 0.30. Furthermore, convergent validity was measured using (1) standardized factor loading (found greater than 0.5), (2) average variance explained (found greater than 0.5), and (3) composite reliability (found greater than 0.7). The confirmatory factor analysis (CFA) of the measurement model indicated the fitness of the constructs (Chi-square minimum [CMIN]=357.62; CMIN/degrees of freedom [DF]=1.80; Goodness of Fit [GFI]=0.90; Adjusted Goodness of Fit Index [AGFI]=0.89; Buntler-Bonett Normed Fit Index [NFI]=0:915; Comparative Fit Index [CFI]=0:93; Root Mean Square Residual [RMR]=0:075; Root Mean Square Error of Approximation [RMSEA]=0:055). Moreover, the structural model fitness was also confirmed (CMIN=488.6; CMIN/DF=1.85; GFI=0.861; AGFI=0.893; NFI=0.987; CFI=0.945; RMR=0:079; RMSEA=0.053). Hence, the results indicated that social media usefulness has a positive and significant effect on health awareness (hypothesis 1: β=.669, *P*<.001), and health awareness has a positive and significant effect on health care education in Pakistan (hypothesis 2: β=.557, *P*<.001).

Conclusions: This study concludes that health care professionals and paramedic staff in private hospitals support the use of social media to raise awareness and provide health care education. It is considered an effective tool for reducing the disease burden in Pakistan. The study results also revealed that young health care professionals are more inclined toward social media usage and express the need for legislation to support it and establish a monitoring process to avoid misinformation.

(JMIR Hum Factors 2025;12:e65745) doi:10.2196/65745

KEYWORDS

social media; health awareness; health education; innovation diffusion theory; structural equation modeling; disease burden; healthcare facilities; health professionals; misinformation; cost effective

Introduction

Background

In Pakistan, the health care system is resource-strapped, leading to challenges in preventing infectious diseases. The country has a high maternal mortality rate (186 deaths per 100,000 live births) and a high infant mortality rate (56 deaths per 1000 live births) [1]. Many complications arise from a lack of awareness and education about prenatal care, safe delivery practices, and postnatal care [2]. Moreover, poor understanding of nutrition has led to widespread malnutrition among children and pregnant

women in Pakistan [1,2]. Malnutrition in children leads to stunted growth, weakened immunity, and impaired cognitive development [2]. The prevalence of stunting in Pakistan is among the highest in the world, affecting approximately 38% of children under 5 years of age [3].

In this context, health awareness and education campaigns potentially lead to behavioral changes and alleviate the burden on health facilities [4]. Public health campaigns, particularly those focused on preventive measures like vaccination, hygiene, and lifestyle changes, have substantially reduced the incidence of various diseases [3,5]. For example, handwashing campaigns alone have reduced respiratory infections by up to 23% [6]. Awareness campaigns aimed at managing chronic conditions, such as diabetes and hypertension, through education on medication adherence and lifestyle modifications have reduced complications and hospital visits [7]. These interventions have been shown to improve health outcomes and reduce health care costs by up to 25% [8].

Among various tools and health interventions, social media has attained the most global attention to enhance health awareness and education to a wider population and discovered its significant impact in addressing health care issues [9,10]. According to social media statistics, social media users worldwide reached 5.04 billion in January 2024, representing 63% of the world's population, with users growing 6% annually [11]. On average, social media users spend 2 hours and 23 minutes daily on social media platforms and see various content [11,12]. Social media tools, such as Facebook (Meta), TikTok (ByteDance), Twitter (subsequently rebranded as X), Instagram (Meta), and YouTube (Google), facilitate health care professionals in educating and influencing perceptions, including networking, sharing ideas, disseminating information, demonstrating, coaching, consulting, and advertising [12,13]. Existing literature in the health care sector has exhibited the transformative potential of social media and its usefulness in spreading health awareness and education [9,10,12,13].

However, the existing literature lacks comprehensive insights into the role of social media in fostering health awareness and education within developing countries, particularly Pakistan. This study addresses this gap by applying Innovation Diffusion Theory to examine the mediating effect of health awareness and empirically testing the impact of social media on health care education among health care professionals.

Therefore, the primary objective of this study is to investigate the role of social media in enhancing health awareness and its subsequent impact on health care education in Pakistan. By addressing the gap in the literature concerning the adoption of social media for health care communication in developing countries, this research aims to provide empirical evidence and practical insights for leveraging digital platforms to improve health outcomes.

This section comprehensively reviews existing literature on social media usefulness, health awareness, health care education, and health care sector challenges in Pakistan.

Social Media Usefulness

The global adoption of social media platforms has led to scalable public health interventions, increased precision and effectiveness, and improved resource mobilization [14]. Social media has encouraged healthy behaviors and enabled informed choices through sharing tips, challenges, and success stories [9]. Health care professionals can provide accurate, evidence-based information as trusted sources [11]. Their involvement is vital due to their expertise, credibility, and ability to reach a wide audience. Campaigns through social media on vaccination, hygiene practices, mental health awareness, and other preventive measures have been helpful in countries like Germany and France [9,10].

Moreover, social media has been effectively used to combat the pandemic and other diseases during the COVID-19 pandemic [15]. The social media data were instrumental in tracking the spread of the COVID-19 virus and disseminating accurate information [16]. Campaigns on platforms such as Facebook and Twitter have been used to promote vaccination, leading to increased uptake in various regions [12]. Tailored messaging on HIV prevention for at-risk populations has shown to be more effective than generalized campaigns [13]. Moreover, existing literature has shown examples of successful public health campaigns, such as:

1. #SmearForSmear Campaign: this social media campaign aimed to raise awareness about cervical cancer screening and successfully increase screening rates by leveraging the reach of platforms such as Twitter and Instagram [17].
2. Truth Initiative: an antismoking campaign that effectively used social media to reduce teen smoking rates through engaging content and interactive platforms [18].

Health Awareness

Health awareness is termed as "Enhancing the ability of individuals to understand and use health information to make informed decisions about their health" [13]. Efforts for health awareness are varied and multifaceted, encompassing a wide range of activities aimed at educating the public about health issues, promoting healthy behaviors, and preventing diseases [16]. Health awareness campaigns lead to early detection and timely intervention, which benefits individuals by lowering their health care expenses and reducing the financial burden on the health care system [18]. Health awareness is generally performed using mass media, social media, posters, billboards, workshops, seminars, local outreach programs, influencers, and public figures [12]. However, social media is most useful in the current era due to its cost-effectiveness and wider audience engagement [17]. Social media allows targeted messaging based on demographics, interests, and online behavior [19]. This ensures that health messages reach the people most likely to benefit from them. Moreover, the effectiveness of health awareness through social media can be measured in a few easy ways, such as the number of impressions, likes, shares, comments, hashtag usage, and so on [18,19]. Hence, medical professionals can analyze the efficacy of their awareness messages. There have been several successful health awareness campaigns conducted through social media that have achieved significant outcomes (Textbox 1).

Textbox 1. Health care awareness campaigns.

- **ThisGirlCan (2015)**
  - Objective: Encourage women of all ages and backgrounds to be more physically active, regardless of their shape, size, or fitness level [20].
  - Execution: The campaign featured real women participating in various physical activities, promoting the message that women should feel confident about being active regardless of societal pressures or body image concerns.
  - Platforms used: Instagram, Twitter, Facebook, and YouTube.
  - Outcomes:
    - Widespread engagement: The hashtag #ThisGirlCan was used millions of times, with significant engagement from women sharing their own stories and photos.
    - Behavioral change: The campaign was credited with encouraging over 2.8 million women in the United Kingdom to become more active.
    - Long-term impact: The campaign continued beyond its initial phase, evolving into a broader movement that still influences public attitudes towards women and fitness.
- **#BellLetsTalk (Ongoing)**
  - Objective: Raise awareness about mental health issues and reduce stigma around mental illness in Canada [21].
  - Execution: For every tweet using the hashtag #BellLetsTalk, every text message sent by Bell customers, and every Facebook video view, Bell Canada donates 5 cents to mental health initiatives.
  - Platforms used: Twitter, Facebook, Instagram, and Snapchat.
  - Outcomes:
    - Record participation: The 2021 campaign saw 159 million interactions, raising nearly CAD $8 million (equivalent US $5.56 million) in a single day.
    - Sustained impact: Since its inception, the campaign has raised more than CAD $121 million (equivalent US $84.5 million) for mental health initiatives.
    - Increased conversations: The campaign has significantly increased public dialogue about mental health in Canada, contributing to reducing stigma and promoting mental wellness.

Health Care Education

Health care education refers to a deeper understanding of specific health topics, including the causes, symptoms, prevention, and treatment of diseases and how to maintain or improve health [20]. It involves having factual information and comprehension of health-related subjects [17]. Health care education encourages regular check-ups, screenings, and vaccinations, leading to early detection of diseases and more effective prevention strategies [19]. For those with chronic conditions such as diabetes, hypertension, or asthma, health care education is essential for managing symptoms, adhering to treatment plans, and avoiding complications [13].

In resource-limited economies, insufficient health care education has been reported among the population, which impedes individuals' ability to understand, access, and apply health information effectively [21]. Moreover, low-income individuals may have less access to health resources, education, and information, which hinders their ability to acquire and apply health education [22]. Addressing these barriers in resource-limited countries requires targeted strategies such as improving health literacy, offering culturally sensitive health care education, increasing access to technology, and combating misinformation [21].

Among these targeted strategies, social media has gained the most attention in the high-income world [20]. Health care education through social media has become an increasingly effective method for reaching diverse audiences with health information. Studies suggest that health care education campaigns on social media can achieve behavioral change success rates ranging from 20% to 40% [16,17]. For instance, social media campaigns promoting COVID-19 vaccination have seen varying success, with some countries reporting a 15% to 25% increase in vaccine uptake attributable to social media efforts [15]. Moreover, some campaigns that promote health services (eg, vaccination drives or mental health counseling) report conversion rates (actual service uptake) of 5% to 20%, depending on the call to action and the ease of access to the service [18]. The existing literature has discussed a few notable health care education campaigns executed through social media and validated their impact. One such campaign for spreading health care education is The Amyotropic Lateral sclerosis (ALS) Ice Bucket Challenge (Textbox 2).

Textbox 2. Health care education campaigns through social media.

1. The ALS Ice Bucket Challenge (2014)

- Overview: Participants dumped buckets of ice water over their heads, shared videos on social media, and nominated others to do the same, all to raise awareness and funds for Amyotrophic Lateral Sclerosis (ALS) research [23].
- Impact:
  - Financial: Raised over US $115 million for the ALS Association in just a few months.
  - Awareness: Dramatically increased global awareness of ALS, with millions participating worldwide.
  - Research Advancement: Funds contributed to significant research advancements, including the discovery of new ALS genes.

Pakistan Context

In Pakistan, the Federal and provincial governments jointly administer 60% of the health care system, with the private sector contributing 30%. Autonomous bodies support the remaining 10 percent [24]. The diverse structure and the country's economic challenges pose unique obstacles to effective health care delivery. The country's health care spending is at 2.95% of its Gross Domestic Product, and the Pakistani government always disregarded the economic survey data and faced criticism from the apex medical organization as it necessitated health allocations in line with global guidelines [25]. To provide quality health care, the Pakistan Medical Association (PMA) expressed the need for 6% of the country's Gross Domestic Product allocation, as recommended by the World Health Organization [17], which the country is unable to meet due to economic challenges [26].

Due to these financial limitations, there are reported disparities in health awareness, with urban areas generally having better access than rural regions [2]. Thus, communicable diseases, such as waterborne diseases and vector-borne diseases, remain a more significant concern in rural regions [1,3]. Moreover, with limited access to maternal health care and family planning services, Pakistan witnesses high maternal and child mortality rates [2]. Diseases such as malaria, tuberculosis, and hepatitis are prevalent, which exacerbates the situation [24]. Moreover, stigma about sexually transmitted infections results in inadequate prevention and treatment efforts that result in a psychological attack on patients [25]. Poor sanitation and hygiene contribute to outbreaks of diseases like cholera and dysentery in Pakistan and many preventable diseases, such as polio and measles, persist due to low vaccination rates [23].

Amid financial crises, health care awareness and education are known to be effective in reducing the disease burden rate [8]. There are several traditional efforts made by the Government of Pakistan Ministry of Health in the past to improve disease awareness and reduce the patient burden [24]. These include the following:

1. Engaging community health workers in disseminating health information, especially in rural areas. However, the number of community health workers to cover the target population is very limited and requires substantial spending [24].
2. Use of radio and television for health awareness programs, public service announcements, and talk shows but requires financial resources [27].
3. Newspapers, magazines, and pamphlets are also used to spread health information, but considering the low literacy rate and cost of print media, these are not found to be much effective for the target population [27].

However, among the modern techniques, social media has gained much attention from health care communities globally [7,10]. Pakistan has also witnessed the widespread of social media but in academics, e-commerce, entertainment, media, politics, sports, and religious sectors. In Pakistan, 29.5% of the population is on social media, whereas 77.8% has active mobile connections, meaning these users also have access to social media platforms [28].

The health care sector is structured to depend on government approval for any initiative [29]. The government has not devised any legislation to promote the use of social media for awareness [24,27]. This mainly restricts the government's health departments and health care professionals from using this tool to publicize health messages. Some efforts have been made to digitalize the health care sector using the eHealth concept and implement a few apps to initiate telehealth and telemedicine [30]. Also, mobile messaging and caller voice tunes are found to be significant in health care services awareness [30]. However, the interpretation and usage of social media tools are still scarce [31]. For this reason, it is imperative to investigate the social media usefulness among health care professionals in Pakistan and to identify whether these health care professionals support the use of social media for health awareness and education.

Theoretical Foundation

A few theories explain the behaviors toward acceptance of any new digitalized tool and address their attitude [32], including the innovation diffusion theory proposed by Rogers in 1962, the Theory of Reasoned Action proposed by Fishbein and Ajzen in 1967, the Self-Efficacy Theory proposed by Bandura in 1977, the Theory of Planned Behavior proposed by Ajzen in 1985, the Social Cognitive Theory proposed by Bandura in 1986, and the Technology Acceptance Model proposed by Davis in 1986 [32]. Among these various behavioral theories, the Innovation Diffusion Theory (IDT) offers a framework for understanding how new ideas, behaviors, or innovations spread within a population [33].

In public health, innovation diffusion theory helps explain how new health interventions, practices, or policies are adopted and disseminated within communities [33]. The theory identifies 5

stages individuals and communities pass when adopting an innovation: knowledge, persuasion, decision, implementation, and confirmation [34]. Figure 1 explains these five stages of the innovation-decision process.

To understand social media acceptance for health awareness and education, the theory has categorized individuals within a population into subgroups based on their readiness to adopt innovations. Innovators and early adopters are likely to adopt innovations early, while most of the population follows suit over time. Laggards are the last to adopt [33]. Figure 2 explains these subgroups of individuals.

Figure 1. Innovation diffusion process.

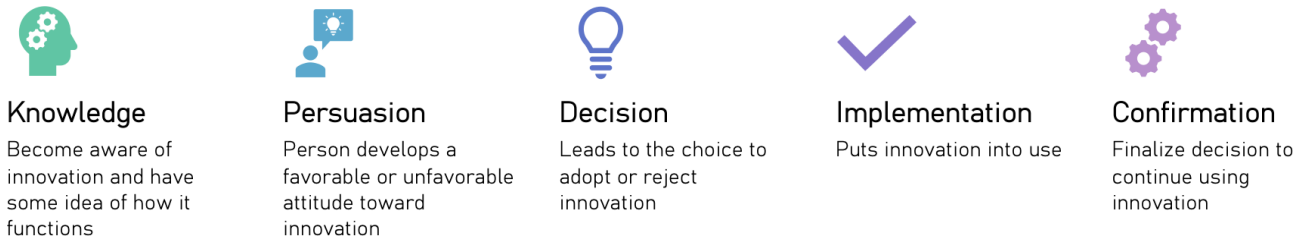
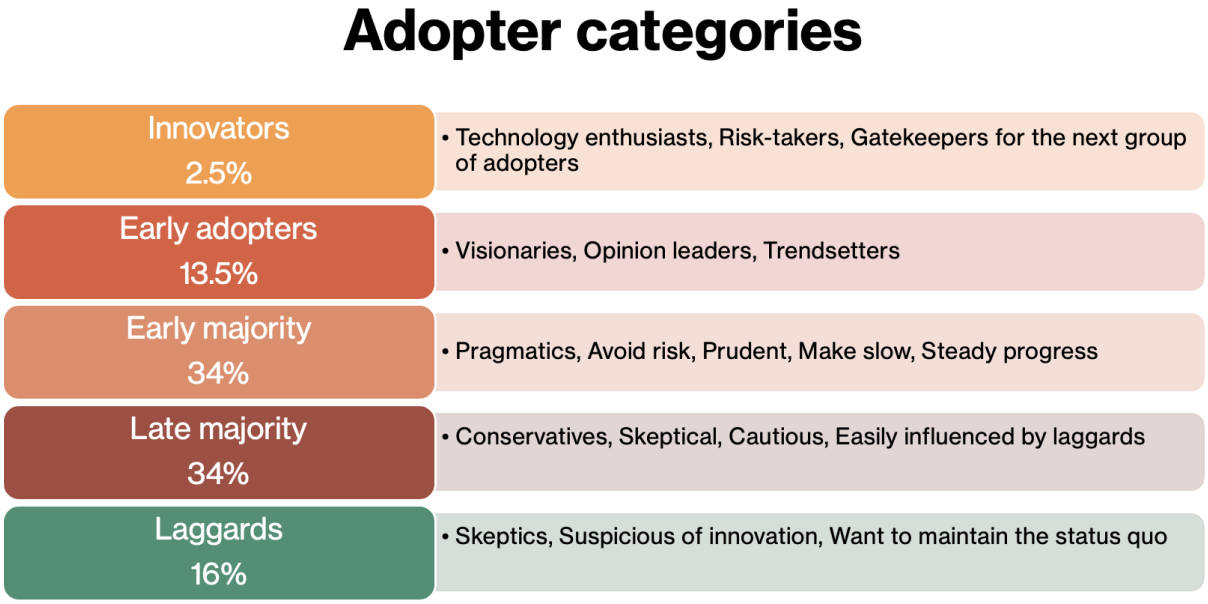


Figure 2. Adopter categories.



In health care education and awareness, the diffusion of innovation is used to fast-track the acceptance of crucial public health digitalized interventions that ordinarily intend to influence the behavior of a social system [34]. Thus, this theory provides the foundation for understanding social media usage and its adoption attitude in a particular community.

In Pakistan, the ineffective innovation management approaches and deprived diffusion of innovation strategies are hurdles to realizing the importance of social media and its usage for health care education [24,25]. The deprived diffusion of innovation is a significant factor in the inability to attain the desired output. Even for digitalized hospital tools other than social media, the country needs more innovative and user-friendly equipment in the hospital sector [25]. The notion of consistent health care challenges and inadequate health resources in Pakistan is also evident from the recent Joint External Evaluation report [24].

For this reason, it is imperative to investigate the impact of social media's effective use on health awareness and health care education in Pakistan using IDT. This may result in highlighting and understanding the effectiveness of social media tools in reducing the disease burden in the country.

Figure 3 explains the theoretical framework of the current study.

Figure 3. Conceptual framework of the study.

Hence, the current study hypotheses are:

1. Hypothesis 1 (H1): social media usefulness positively and significantly impacts health awareness in Pakistan.
2. Hypothesis 2 (H2): health awareness has a significant positive impact on health care education in Pakistan.

Methods

Instrument Development

A quantitative technique was applied to test the study hypothesis and validate the proposed model using the structural equation modeling (SEM) technique [35]. SEM is deemed the most suitable due to its ability to simultaneously test complex relationships among multiple variables and validate the proposed conceptual model. Data analysis was carried out using AMOS 26.0 software (IBM Corp) [36]. Since the study aims to identify social media's impact on health awareness leading to health care education and uses the IDT as a theoretical foundation, scales from previous literature have been adopted. The main advantage of adopting these scales is that they have already been validated in the health industry's acceptance behavior of new technological tools. For instance, the measurement scale for Innovative Technology (social media) usefulness was adopted from a previous study [37], and measurement scales for health awareness have been adopted from another study [38]. Aligning with the psychometric literature, 4-item scales for each construct based on their higher loadings have been adopted [37]. The modified 8-item scale for measuring health care education was adopted based on the questionnaire of Ho et al [39]. For the measurement of each item, a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree) was used [38].

Data Collection

A questionnaire was designed based on the variables of the present study, utilizing the aforementioned scales, which are provided in Multimedia Appendix 1. A personally administered questionnaire was used for data collection [32]. For data collection, the current research to eliminate targeted health care professionals and paramedic staff of 2 private hospitals in Lahore, Pakistan. The questionnaire was distributed to more than 450 health care professionals and paramedic staff using convenience sampling and requested their valuable feedback. However, only 389 useful responses were received.

Ethical Considerations

As a matter of ethical concerns, prior permission from the hospital management was obtained, and the purpose of the study was explained to all respondents through "informed consent." It was assured that respondents' privacy would be ensured through confidentiality. Ethical approval was not required for this study as it did not involve identifiable personal data, or interventions falling under the scope of institutional or national ethics review requirements. In accordance with the Declaration of Helsinki (World Medical Association, 2013) research ethics policy, studies that do not involve sensitive data are exempt from formal ethics board approval.

Results

Primary Data Screening

Initial screening was carried out among 389 useful responses to eliminate unusual responses, mainly incomplete questionnaires and respondents who marked similar scores on all items. Furthermore, questionnaires with more than 4 missing values were also excluded. The remaining missing values were replaced with the series mean/average for the rest of the responses. Therefore, the final useful sample was reduced to 340.

Table 1 provides the demographics of the study respondents. Among 340 respondents, 58% (197/340) were male, and the rest were female (143/340, 42%). A total of 91% (309/340) of the respondents were between 25 and 35 years of age, which may lead to the fact that most health care professionals and paramedic staff in private hospitals in Lahore are young. However, we may be unable to neglect the 6% of the population under 25 years old and more eager to start their career in the health care sector. The remaining 3% were more than 35 years of age and had extensive health industry work experience.

Table . Sociodemographic characteristics of study respondents (N=340).

Characteristics	Values, n (%)
Sex	
Male	197 (58)
Female	143 (42)
Age (years)	
<25	21 (6)
25-35	309 (91)
>35	10 (3)
Social media experience (years)	
<1	34 (10)
2-5	207 (61)
>5	99 (29)
Monthly income[a]	
Less than PKR[b] 40,000 (US $142.98)	101 (30)
PKR 40,000-80,000 (US $142.98-US $285.97)	197 (58)
More than 80,000 (US $285.97)	42 (12)
Education	
Less than secondary school	63 (18)
Secondary school to graduation	259 (76)
Medical (MBBS[c]/BDS)[d] or Masters degree	18 (6)
Health services experience (years)	
<3	21 (6)
3-6	225 (66)
>6	94 (28)

[a]A currency exchange rate of US $1=280 PKR is applicable.

[b]PKR: Pakistani rupees.

[c]MBBS: Bachelor of Medicine, Bachelor of Surgery.

[d]BDS: Bachelor of Dental Surgery.

Data Analysis

An exploratory factor analysis was initially performed on the 16 measurement items to confirm the underlying relationships [36]. For sample adequacy, the KMO value (0.850) was greater than the recommended value (0.60), and the significance value of *P*0.001 confirmed the Bartlett test of sphericity, as shown in Table S1 in Multimedia Appendix 2. All those factors were retained for data analysis, with factor loading greater than 0.50 and Eigenvalue greater than 1.0. Hence, this technique provided 3 factors that explained 72.2% of the variance after varimax rotation. Furthermore, scale reliability coefficients were greater than the acceptable value (0.70) [40], and Shapiro-Wilk tests were conducted to check the data normality [40], as shown in Table S2 in Multimedia Appendix 2 . The results of both tests were significant, meaning there was no normality issue in the data.

For further data analysis, a two-stage structural equation modeling (SEM) technique was used [36-38]. This approach allows researchers to evaluate the measurement and structural models separately using two dissimilar subsamples. For this reason, a sample of 340 participants was divided into two parts: a sample of 170 individuals was used for the measurement model, and a similar sample size was used for the structural model assessment to attain impartial results.

Reliability and Validity Measures

According to Hair et al [35], finding the reliability of each item and construct in the research study is essential. Therefore, squared multiple correlation (SMC) was used to find the reliability of each reliability item of each measurement item. SMC represents “the amount of variance explained by an individual indicator/construct of its respective factor and measured by the square of its (indicator’s) standardized factor loading” [40].

Hair [35] indicated that the cutoff value of SMC is 0.30. Hence, it is evident from Table 2 that all the SMC values are greater than 0.30. For measuring the reliability of each variable, the

Cronbach alpha value was used. As suggested in the literature, the cutoff value of Cronbach alpha is 0.70 [36,37]; it is evident from Table 2 that all variables' Cronbach alpha values are more than 0.70. Furthermore, Hair et al [35] also highlighted that for measuring the convergent validity, there are three common approaches: (1) standardized factor loading (0.5 or greater), (2) average variance explained (0.5 or higher), and (3) composite reliability (0.7 or above).

Table . Measurement of reliability and validity.

Constructs and item	Values, mean (SD)	Standardized factor loading	Squared multiple correlation	Composite reliability	Cronbach alpha	Average variance explained
Social media usefulness (SMU)						
SMU1	5.19 (1.65)	0.781	0.607	0.788	0.774	0.589
SMU2	5.33 (1.71)	0.780	0.687			
SMU3	5.01 (1.12)	0.777	0.671			
SMU4	5.05 (1.32)	0.787	0.646			
Health awareness (HA)						
HA1	5.45 (1.56)	0.859	0.677	0.769	0.773	0.592
HA2	5.98 (1.45)	0.874	0.681			
HA3	5.39 (1.43)	0.881	0.675			
HA4	5.51 (1.39)	0.867	0.654			
Health care education (HED)						
HED1	5.23 (1.77)	0.719	0.651	0.783	0.789	0.678
HED2	5.22 (1.74)	0.775	0.629			
HED3	5.60 (1.64)	0.721	0.622			
HED4	5.41 (1.49)	0.739	0.637			
HED5	5.22 (1.15)	0.787	0.663			
HED6	5.11 (1.23)	0.772	0.654			
HED7	5.09 (1.43)	0.759	0.691			
HED8	5.32 (1.59)	0.768	0.609			

Hence, each standardized factor loading was statistically significant (P<.001), and values ranged from 0.719 to 0.881, thus validating adequate convergent validity as shown in Table 2. In addition, the average value explained (AVE) values and construct reliability were also more than their cut-off level of 0.5 and 0.7, respectively, and were statistically significant. Therefore, these measures confirm sufficient convergent validity.

A comparison of shared variance between factors with the average variance explained by individual factors was ensured for measuring the discriminant validity. The diagonal value should be greater than the non-diagonal value to confirm adequate discriminant validity [40]. Hence, results given in Table S3 in Multimedia Appendix 2indicates the correlation matrix of constructs, where non-diagonal elements are correlated among constructs and diagonal elements are the square root of average variance explained (AVE) by that construct and clearly explain that all three constructs differ.

Measurement Model

Using statistical software AMOS 26.0, the measurement model's confirmatory factor analysis (CFA) was performed [40,41]. This provided a passable model fit for the primary measurement model (CMIN=1220.23; CMIN/DF=1.98; GFI=0.72; AGFI=0.80; NFI=0.81; CFI=0.826; RMR=0.089; RMSEA=0.082); however, the number of indicators per item was large; for instance, the number of indicators for health care education were 8. Consequently, 2 items were deleted for further refinement to obtain an adequate model fit through the CFA of the measurement model. This refinement was conducted by deleting items one by one, based on their standardized residual; that is, that item was first deleted, which had a larger error variance than their measurement items. Each item was carefully reviewed before deleting it to ensure that, from a theoretical viewpoint, its error variance also seemed rational. The refinement and assessment process for every construct was first evident by Churchill [41]. Churchill defined this process as; "Though this application may be satisfactory during the early stages of research on a construct, the use of factor analysis in a confirmatory fashion would seem better at later stages." Furthermore, Gerbing and Anderson's [42] study also provided support to Churchill's argument and stated that; "to demonstrate that an explicit evaluation of Unidimensionality is accomplished with a confirmatory factor analysis of the individual measures as specified by a multiple-indicator measurement model. Coefficient alpha is important in assessing reliability but does not assess dimensionality. Although item-total correlations and exploratory factor analysis can provide useful preliminary

analyses, particularly in the absence of a sufficiently detailed theory, they do not directly assess unidimensionality. The reason is that a confirmatory factor analysis assesses the internal consistency and external consistency criteria of unidimensionality implied by the multiple-indicator measurement model.

Hence, this refinement process provided an adequate model fit (CMIN=357.62; the ratio of χ^2 to degree of freedom value (1.80) is remarkably less than its recommended value (5.0). CMIN/DF=1.80; GFI=0.90; AGFI=0.89; NFI=0:915; CFI=0:93; RMR=0:075; RMSEA=0:055) as shown in Table 3.

Table . Structural equation modeling fit indices for the confirmatory factor analysis model.

Fit indices	Cut-off criteria	Results obtained
Absolute fit indices		
Chi-square (*df*)	N/A[a]	357.62 (170)
χ^2/df (CMIN/DF)	<5.00	1.800
Root Mean Square Error of Approximation	<0.06	0.055
Goodness of Fit Index	>0.85	0.900
Adjusted Goodness of Fit Index	>0.85	0.890
Incremental fit indices		
Buntler-Bonett Normed Fit Index	>0.90	0.915
Comparative Fit Index	>0.93	0.930
Tucker Lewis Index	>0.90	0.941
Incremental Fit Index	>0.90	0.932
Parsimonious fit indices		
Parsimony Goodness-Fit Index	>0.50	0.798
Parsimony Normed Fit Index	>0.50	0.848

[a]N/A: not applicable.

Structural Model

The research hypotheses were tested through structural model estimation [40]. Therefore, a second subsample (n=170) was used and provided the adequate structural model fit (CMIN=488.6; CMIN/DF=1.85; GFI=0.861; AGFI=0.893; NFI=0.987; CFI=0.945; RMR=0:079; RMSEA=0.053) as shown in Table 4.

Table . Hypothesized structural model fit indices.

Fit index	Cut-off criteria	Results obtained
Absolute fit indices		
Chi-square (*df*)	N/A[a]	488.6 (170)
χ^2/df (CMIN/DF)	<5.00	1.850
Root Mean Square Error of Approximation	<0.06	0.053
Goodness of Fit Index	>0.85	0.861
Adjusted Goodness of Fit Index	>0.85	0.893
Incremental fit indices		
Buntler-Bonett Normed Fit Index	>0.90	0.987
Comparative Fit Index	>0.93	0.945
Tucker Lewis Index	>0.90	0.976
Incremental Fit Index	>0.90	0.943
Parsimonious fit indices		
Parsimony Goodness-Fit Index	>0.50	0.680
Parsimony Normed Fit Index	>0.50	0.609

[a]N/A: not applicable.

Moreover, the significance of the hypothesis is shown in Table 5. Testing of H1 revealed that social media usefulness has a significant effect on health awareness (H1: β=.669, *P*<.001), supporting the IDT proposed by Rogers. Testing of H2 also supplied similar significant outcomes to the current research model (H2: β=.557, *P*<.001), confirming the findins of prior work.

Table . Hypothesis testing results.

S.no.	Impact of	Impact on	Hypothesis	Path coefficient	*P* value	Outcome
1	SMU[a]	HA[b]	H1[c]	0.669[d]	.001	Significant impact
2	HA	HED[e]	H2[f]	0.557[d]	.001	Significant impact

[a]SMU: social media usefulness.

[b]HA: health awareness.

[c]H1: hypothesis 1.

[d]*P*<.001.

[e]HED: health care education.

[f]H2: hypothesis 2.

Discussion

Principal Findings

This study examined the impact of social media on health awareness and its role in health care education, using IDT as the theoretical framework in Pakistan. A conceptual model was developed to assess the mediating effect of health awareness on the relationship between social media usage and health care education. The hypothesis testing results demonstrated significant relationships, with social media positively influencing health awareness (H_1: path coefficient=0.669, *P*<.001) and health awareness positively impacting health care education (H_2: path coefficient=0.557, *P*<.001). These findings highlight the potential of social media as an effective tool for promoting health care education when strategically leveraged by health care stakeholders.

The study also revealed that younger and more educated health staff are more inclined to use social media, while older groups, though recognizing its usefulness, engage less actively. This demographic insight underlines the need for tailored engagement strategies to maximize participation. Limited research in developing countries, including Pakistan, has tested the application of IDT to health care professionals' adoption of social media for health care education. The standardized direct effects from hypothesis testing affirm the significant relationship between social media usefulness, health awareness, and health care education in the Pakistani context.

Given that developing countries are often classified as late adopters of technology, according to IDT, the findings emphasize the importance of region-specific social media content on platforms, such as Facebook, WhatsApp, or locally developed alternatives, for health care interventions. IDT further suggests that simplifying innovations can facilitate broader adoption, making it easier for health care professionals and the general population to engage with digital health initiatives. These insights highlight the need for cost-effective, user-centric solutions to enhance healthcare outcomes in Pakistan.

The government of Pakistan can also play a significant role in spreading health awareness. They can use social media to run campaigns on maternal health, child nutrition, and disease prevention. The health care professionals in private hospitals also indicated that live sessions and webinars on platforms like Facebook Live and Instagram Live to educate the patients and hospital staff on various health issues and answer their real-time questions have been very useful.

Using social media for health care education in Pakistan could be a game-changer, surpassing the expectations of spreading awareness regarding health needs in a particular community. This potential is not just theoretical; it has been proven in developed countries, such as the US, Jordan, the United Kingdom, and Europe, where social media was used during the COVID-19 pandemic to combat the pandemic and other notable diseases [15]. Moreover, social media platforms have been used to promote telemedicine services, enabling health care professionals to connect with patients remotely for consultations, follow-ups, and monitoring of chronic conditions [16,30].

Overall, this study validates the use of social media to improve public health care education through public awareness. Therefore, the study concludes that health care professionals should use social media tools to inform the wider public and address healthcare issues in Pakistan.

Conclusion

The findings of this study underscore the significant role that social media plays in enhancing health care awareness and education among various populations in developing countries like Pakistan. With their wide reach and interactive features, social media platforms have proven to be effective tools for disseminating health-related information and engaging individuals in meaningful health conversations. Social media breaks down geographical barriers, allowing health information to reach a global audience instantly. This broadens the scope of health care awareness campaigns and makes information accessible to a diverse demographic, including those in remote or underserved areas. The interactive nature of social media facilitates active engagement between health care providers and the public. This interaction fosters a sense of community, encourages the sharing of personal health experiences, and allows for immediate feedback and clarification of health information. Also, social media platforms enable the

dissemination of personalized health information tailored to specific audiences' needs. Moreover, the real-time nature of these platforms ensures that information is up to date, which is crucial during health emergencies or outbreaks.

Health care education through social media can help understand how to prevent common diseases such as malaria, tuberculosis, hepatitis, and polio through vaccination, sanitation, and hygiene in Pakistan. Awareness campaigns can encourage healthier lifestyles, reducing the incidence of chronic diseases such as diabetes, hypertension, and heart disease. Educating women about prenatal and postnatal care, safe delivery practices, and child nutrition can significantly lower maternal and infant mortality rates. Awareness through social media about proper nutrition, immunization, and hygiene practices helps improve children's health and development in the rural outskirts of Pakistan. Also, awareness about communicable diseases and their transmission can help in preventing outbreaks of diseases like HIV/AIDS, tuberculosis, and waterborne illnesses.

Governmental policies provide a framework for regulating the content shared on social media, ensuring that the information disseminated is accurate, reliable, and in line with public health guidelines. Information shared by government-endorsed social media accounts is more likely to be trusted by the public. Official policies lend credibility and legitimacy to health messages, increasing the likelihood of public acceptance and compliance. All active organizations (govt and private) working in the health industry of Pakistan seek prior permission or ethical clearance to initiate any new work. Without government support, the organizations lack the confidence to spread and advertise any health-related content due to fear that they may be subject to local resistance.

Despite its benefits, the use of social media in health care awareness also presents challenges, such as the spread of misinformation and the need for privacy and data protection. Health care organizations must implement strategies to verify information and ensure the credibility of the content shared. Hence, the study suggests that Pakistan's health care organizations should continue leveraging social media to enhance health communication strategies and education. Ongoing research is also needed to explore new ways to maximize social media's benefits while mitigating its risks.

In conclusion, social media is a powerful tool for health care awareness, offering unprecedented opportunities to reach and engage with a wide audience. By harnessing its potential in developing economies, health care organizations can improve public health literacy, promote healthy behaviors, and ultimately contribute to better health outcomes. Collaborative efforts between health care professionals, social media platforms, and policymakers will be crucial in leveraging social media's full potential for public health advancement.

Future Research

Future research should move beyond compliance with existing social media policies to explore strategies and variables that demonstrate tangible improvements in the health care sector. This could include investigating how social media can influence enhanced patient outcomes through better communication and engagement, broaden health awareness campaigns by targeting diverse and underserved populations, and improve access to health care services through innovative digital solutions. Additionally, longitudinal studies examining the long-term impact of social media interventions on public health indicators would provide valuable insights. Identifying and analyzing these variables will contribute to a more comprehensive understanding of social media's transformative potential in health care beyond mere policy adherence.

Acknowledgments

The authors would like to thank Mr Nabil Ahmed, Director at Integral Global Consulting LLC, Tucker, GA, US for providing the necessary resources and facilities that enabled the successful execution of this research. The authors also want to extend their special thanks to all the respondents and hospital management in Lahore, Pakistan, for their feedback and the collaborative environment that fostered productive discussions and the exchange of ideas.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire.
[DOCX File, 19 KB - humanfactors_v12i1e65745_app1.docx]

Multimedia Appendix 2
Additional tables.
[DOCX File, 17 KB - humanfactors_v12i1e65745_app2.docx]

References

1. Shaeen SK, Tharwani ZH, Bilal W, Islam Z, Essar MY. Maternal mortality in Pakistan: challenges, efforts, and recommendations. Ann Med Surg (Lond) 2022 Sep;81:104380. [doi: 10.1016/j.amsu.2022.104380] [Medline: 36042926]

2. Hanif M, Khalid S, Rasul A, Mahmood K. Maternal mortality in rural areas of Pakistan: challenges and prospects. Rural Health 2021;7. [doi: 10.5772/intechopen.96934]
3. Soofi SB, Khan A, Kureishy S, et al. Determinants of stunting among children under five in Pakistan. Nutrients 2023 Aug 7;15(15):3480. [doi: 10.3390/nu15153480] [Medline: 37571417]
4. Bul K, Holliday N, Bhuiyan MRA, Clark CCT, Allen J, Wark PA. Usability and preliminary efficacy of an artificial intelligence-driven platform supporting dietary management in diabetes: mixed methods study. JMIR Hum Factors 2023 Aug 9;10:e43959. [doi: 10.2196/43959] [Medline: 37556198]
5. Siddiqi DA, Ali RF, Shah MT, et al. Evaluation of a mobile-based immunization decision support system for scheduling age-appropriate vaccine schedules for children younger than 2 years in Pakistan and Bangladesh: lessons from a multisite, mixed methods study. JMIR Pediatr Parent 2023 Feb 17;6:e40269. [doi: 10.2196/40269] [Medline: 36800221]
6. Mo Y, Pham TM, Lim C, et al. The effect of hand hygiene frequency on reducing acute respiratory infections in the community: a meta-analysis. Epidemiol Infect 2022;150:e79. [doi: 10.1017/S0950268822000516]
7. Eaton MC, Probst YC, Smith MA. Characterizing the discourse of popular diets to describe information dispersal and identify leading voices, interaction, and themes of mental health: social network analysis. JMIR Infodemiology 2023 May 5;3:e38245. [doi: 10.2196/38245] [Medline: 37159259]
8. Allegranzi B, Tartari E, Pittet D. 'Seconds SAVE LIVES – clean your hands': the 5 May 2021 World Health Organization's SAVE LIVES: Clean Your Hands campaign. Int J Infect Control 2021;17(1):21418. [doi: 10.3396/ijic.v17.21418]
9. Subramanyam C, Becker A, Rizzo J, Afzal N, Nong Y, Sivamani R. Visibility of board-certified dermatologists on TikTok. JMIR Dermatol 2024 Jan 5;7:e46085. [doi: 10.2196/46085] [Medline: 38180786]
10. Griffis HM, Kilaru AS, Werner RM, et al. Use of social media across US hospitals: descriptive analysis of adoption and utilization. J Med Internet Res 2014 Nov 27;16(11):e264. [doi: 10.2196/jmir.3758] [Medline: 25431831]
11. Global social media statistics. DATAREPORTAL. URL: https://datareportal.com/social-media-users [accessed 2025-03-25]
12. Terrasse M, Gorin M, Sisti D. Social media, e-health, and medical ethics. Hastings Cent Rep 2019 Jan;49(1):24-33. [doi: 10.1002/hast.975] [Medline: 30790306]
13. Smith MA. NodeXL: simple network analysis for social media. Presented at: Proceedings of the 2013 International Conference on Collaboration Technologies and Systems (CTS); May 20-24, 2013; San Diego, CA, USA URL: https://ieeexplore.ieee.org/document/6567211
14. Bazzano AN, Patel T, Nauman E, Cernigliaro D, Shi L. Optimizing telehealth for diabetes management in the deep south of the United States: qualitative study of barriers and facilitators on the patient and clinician journey. J Med Internet Res 2024 Jan 1;26:e43583. [doi: 10.2196/43583] [Medline: 37976468]
15. Balki E, Holland C, Hayes N. Use and acceptance of digital communication technology by older adults for social connectedness during the COVID-19 pandemic: mixed methods study. J Med Internet Res 2023 Aug 2;25:e41535. [doi: 10.2196/41535] [Medline: 37531187]
16. Santos DS, Batistelli CRS, Lara MMDS, Ferreira EDS, Moreira TR, Cotta RMM. The effectiveness of the use of telehealth programs in the care of individuals with hypertension and, or diabetes mellitus: systematic review and meta-analysis. Diabetol Metab Syndr 2022 May 28;14(1):76. [doi: 10.1186/s13098-022-00846-5] [Medline: 35643495]
17. Lenoir P, Moulahi B, Azé J, Bringay S, Mercier G, Carbonnel F. Raising awareness about cervical cancer using Twitter: content analysis of the 2015 #SmearForSmear campaign. J Med Internet Res 2017 Oct 16;19(10):e344. [doi: 10.2196/jmir.8421] [Medline: 29038096]
18. Niederdeppe J, Farrelly MC, Haviland ML. Confirming "truth": more evidence of a successful tobacco countermarketing campaign in Florida. Am J Public Health 2004 Feb;94(2):255-257. [doi: 10.2105/ajph.94.2.255] [Medline: 14759936]
19. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40. [doi: 10.2196/jmir.2952] [Medline: 24550083]
20. Bauman A, McNeil N, Nicholson M, et al. Impact of the first year of the "This girl can" physical activity and sport mass media campaign in Australia. BMC Public Health 2023 Feb 15;23(1):333. [doi: 10.1186/s12889-023-15091-2] [Medline: 36793043]
21. Booth RG, Allen BN, Bray Jenkyn KM, Li L, Shariff SZ. Youth mental health services utilization rates after a large-scale social media campaign: population-based interrupted time-series analysis. JMIR Ment Health 2018 Apr 6;5(2):e27. [doi: 10.2196/mental.8808] [Medline: 29625954]
22. Stronks K, van de Mheen HD, Mackenbach JP. A higher prevalence of health problems in low income groups: does it reflect relative deprivation? J Epidemiol Community Health 1998 Sep;52(9):548-557. [doi: 10.1136/jech.52.9.548] [Medline: 10320855]
23. Song P. The Ice Bucket Challenge: the public sector should get ready to promptly promote the sustained development of a system of medical care for and research into rare diseases. Intractable Rare Dis Res 2014 Aug;3(3):94-96. [doi: 10.5582/irdr.2014.01015] [Medline: 25364651]
24. Khan SJ, Asif M, Aslam S, Khan WJ, Hamza SA. Pakistan's healthcare system:a review of major challenges and the first comprehensive universal health coverage initiative. Cureus 2023 Sep;15(9):e44641. [doi: 10.7759/cureus.44641] [Medline: 37799252]

25. Naz F, Asghar K. Patients satisfaction and healthcare delivery system in public sector hospitals. Pak Soc Sci Rev 2023 Jul;7(3):352-362. [doi: 10.35484/pssr.2023(7-III)28]
26. PMA expresses grave concern over deteriorating health indicators. BOL News. URL: https://www.pmacentre.org.pk/new/Newscuttings%2010-06-2023%20-%20Health%20Indicators.pdf [accessed 2025-03-25]
27. Saleem A, Rizvi WR, Farrukh M, Saleem M. Influence of radio programs on development of small & medium enterprises in central Punjab Pakistan. PalArch's J Archaeol Egypt/ Egyptol 2021;18(10):116-140.
28. Global Social Media Statistics. Pakistan. URL: https://datareportal.com/reports/digital-2024pakistan#:~:text=Pakistan%20was%20home%20to%2071.70,percent%20of%20the%20total%20population [accessed 2025-03-25]
29. Shaikh BT, Ali N. Universal health coverage in Pakistan: is the health system geared up to take on the challenge? Global Health 2023 Jan 12;19(1):4. [doi: 10.1186/s12992-023-00904-1] [Medline: 36635698]
30. Bilal W, Qamar K, Siddiqui A, Kumar P, Essar MY. Digital health and telemedicine in Pakistan: Improving maternal healthcare. Ann Med Surg (Lond) 2022 Sep;81:104425. [doi: 10.1016/j.amsu.2022.104425] [Medline: 36046714]
31. Ittefaq M, Seo H, Abwao M, Baines A. Social media use for health, cultural characteristics, and demographics: a survey of Pakistani millennials. Digit Health 2022;8:20552076221089454. [doi: 10.1177/20552076221089454] [Medline: 35401998]
32. Ho KF, Chang PC, Kurniasari MD, Susanty S, Chung MH. Determining factors affecting nurses' acceptance of a care plan system using a modified technology acceptance model 3: structural equation model with cross-sectional data. JMIR Med Inform 2020 May 5;8(5):e15686. [doi: 10.2196/15686] [Medline: 32369033]
33. Dearing JW. Applying diffusion of innovation theory to intervention development. Res Soc Work Pract 2009 Sep 1;19(5):503-518. [doi: 10.1177/1049731509335569] [Medline: 20976022]
34. Zhang X, Yu P, Yan J, Ton A M Spil I. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15:71. [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
35. Hair J, Hult G, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM): Sage Publications; 2013.
36. Verdam MGE, Oort FJ, Sprangers MAG. Using structural equation modeling to investigate change and response shift in patient-reported outcomes: practical considerations and recommendations. Qual Life Res 2021 May;30(5):1293-1304. [doi: 10.1007/s11136-020-02742-9] [Medline: 33550541]
37. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
38. Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Information Systems Research 2000 Dec;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]
39. Ho KF, Ho CH, Chung MH. Theoretical integration of user satisfaction and technology acceptance of the nursing process information system. PLoS One 2019;14(6):e0217622. [doi: 10.1371/journal.pone.0217622] [Medline: 31163076]
40. Hair J. Multivariate Data Analysis: Pearson Prentice Hall; 2006. URL: https://drnishikantjha.com/papersCollection/Multivariate%20Data%20Analysis.pdf [accessed 2024-08-21]
41. Churchill GA. A paradigm for developing better measures of marketing constructs. Journal of Marketing Research 1979 Feb;16(1):64-73. [doi: 10.1177/002224377901600110]
42. Anderson JC, Gerbing DW. The meffect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. Psychometrika 1984 Jun;49(2):155-173. [doi: 10.1007/BF02294170]

Abbreviations

AGFI: Adjusted Goodness of Fit Index
AVE: average value explained
CFA: confirmatory factor analysis
CFI: Comparative Fit Index
CMIN: chi-square minimum
DF: degrees of freedom
GFI: Goodness of Fit
H1: hypothesis 1
H2: hypothesis 2
IDT: Innovation Diffusion Theory
NFI: Buntler-Bonett Normed Fit Index
PMA: Pakistan Medical Association
RMR: Root Mean Square Residual
RMSEA: Root Mean Square Error of Approximation
SEM: structural equation modeling
SMC: squared multiple correlation

Edited by A Kushniruk; submitted 24.08.24; peer-reviewed by A Bilal, M Khairulbahri, M Javed; revised version received 09.02.25; accepted 10.02.25; published 07.04.25.

Please cite as:
Munir MM, Ahmed N
Using Social Media Platforms to Raise Health Awareness and Increase Health Education in Pakistan: Structural Equation Modeling Analysis and Questionnaire Study
JMIR Hum Factors 2025;12:e65745
URL: https://humanfactors.jmir.org/2025/1/e65745
doi:10.2196/65745

Patients' and Physicians' Experience With and Acceptability of a Telemedicine Cabin: Mixed Methods Study

Caroline Villela Galvão de França*, MD; Paola Boaro Segalla*, MD; Felipe Sebastião de Assis Reis*, MD, MSc; José Ricardo Silveira Pereira*, MBA; Alexandre Oliveira de Mattos*, BSc; Roberta de Moura Ferron*, BSc; Cleyton Zanardo de Oliveira*, MSc; Jéssica Bassani Borges*, MBA, PhD; Lilian Quintal Hoffmann*, MBA; Edmundo Di Giaimo Caboclo*, MD

Department Occupational Medicine, Hospital BP (A Beneficência Portuguesa de São Paulo), Rua Maestro Cardin, 769, São Paulo, Brazil

*all authors contributed equally

Corresponding Author:
Edmundo Di Giaimo Caboclo, MD
Department Occupational Medicine, Hospital BP (A Beneficência Portuguesa de São Paulo), Rua Maestro Cardin, 769, São Paulo, Brazil

Abstract

Background: Telemedicine represents an essential tool with the potential to reduce health costs, thus avoiding patient displacement and improving patient care outcomes, positioning it as a significant social technology.

Objective: This study aims to analyze the implementation of a telehealth cabin at BP Hospital (A Beneficência Portuguesa de São Paulo), focusing on the evaluation of the experiences of both patients and health care professionals, as well as the acceptability of this tool.

Methods: A mixed methods study was conducted with 229 participants, divided into 2 phases. The first phase involved 40 apparently healthy individuals to assess the usability, experience, and satisfaction of this group for the later safe application in the group with clinical complaints. The second phase included 189 participants, with complaints to assess the usability, experience, and satisfaction of patients and doctors. In both phases, participants completed screening questionnaires (to assess the eligibility criteria), a socioeconomic demographic questionnaire before using the cabin, and a questionnaire including the System Usability Scale and the Net Promoter Score (NPS) after using the cabin.

Results: The data analysis of the first phase showed high acceptance of the telehealth cabin, which supported the progression to the second phase. In the second phase, a high usability score was observed among participants with clinical complaints (mean System Usability Scale score of 85.97, SD 15.50) and a high favorability rating (NPS score of 9.4). Health care professionals also reported favorable results, with a usability score of 67.8 and an NPS of 8.0.

Conclusions: The results of this study reinforce the potential for scaling up this practice based on usability outcomes, and highlight its relevance for the development of public policies aimed at expanding access to quality health care in Brazil. This approach improves the interaction of patients with the health care system, while providing professionals with an extended view of clinical conditions through integrated devices, particularly in areas with limited access to medical care.

(JMIR Hum Factors 2025;12:e55430) doi:10.2196/55430

KEYWORDS

telemedicine cabin; telehealth; teleservice; e-health; connected offices

Introduction

Health is a fundamental right of the population and must be guaranteed to everyone. It is even included in the Universal Declaration of Human Rights of 1948, in Article 25, which defines that everyone has the right to a standard of living that ensures health, well-being, and health care [1]. Although access to health is often presented as a goal in the health policy, there are several challenges to ensure that right into the Brazilian reality, such as inequalities related to the access and use of health services, discontinuities in the geographical distribution of services, especially of medium and high complexity, in addition to the articulation between health systems of several care levels [2,3].

Among the various proposals developed to improve health care at the national level, the use of technological and telecommunication resources for the exchange of information across different levels of health care, between health professionals, and between doctors and patients, has gained prominence.

According to the Ministry of Health, telehealth uses information and communication technologies to promote the expansion and improvement of medical services. Telemedicine is included in telehealth, related only to remote care through technology. According to the World Health Organization, telemedicine is defined as the "delivery of health services where distance to health professionals is a critical factor, through the use of information and communication technologies to the exchange of valid information for the diagnosis, treatment, and prevention of disease and injuries, research and assessment, and to the continuing education of the providers of healthcare and patients."

Telehealth has been used in Brazil since the 1990s, emerging in a decentralized and fragmented way in the health sites [4]. Telemedicine has the potential to reduce health costs, avoid patient displacement, and even improve patient care outcomes [5-7]. Although many studies are being published, especially as a consequence of the COVID-19 pandemic, the implementation of telemedicine can still be challenging, mainly due to technological barriers and poor computer knowledge, and even due to resisting change and the patient's education level [8]. Despite the increasing interest in telehealth, a significant part of the research focuses on the technological aspects rather than on the tool's adoption and acceptance.

One of the limitations of telemedicine is the lack of possibility to monitor the patient throughout medical care, and some telemedicine platform initiatives have been developed for this purpose. However, today, they provide more complete solutions through a digital health ecosystem, which offers medical instruments to be used by the patient, such as a temperature sensor, stethoscope, dermatoscope, oximeter, otoscope, sphygmomanometer, scale, and height sensor. These new functionalities aim to create resources that can improve even more access and affect the quality of life by bringing doctors and patients closer together.

This research aimed to assess the human aspects of the experience and usability when using telehealth cabins with built-in medical devices, thus expanding the diagnostic and interventional capacity of health professionals in clinical, diagnostic, and interventional capacity practice in its remote format.

To this end, the socioeconomic and demographic profile of the population, the experience and usability of the population in the cabin, the doctor's experience in providing care through telehealth cabins, and compliance in handling the equipment available in the self-examination cabin were evaluated.

Methods

Recruitment

This is a mixed methods study using (1) the System Usability Scale (SUS), (2) the Net Promoter Score (NPS), and (3) a thematic analysis based on participant' perceptions. The aim was to assess the usability and favorability of health care provided in a connected cabin for symptomatic and asymptomatic employees of the BP Hospital (Beneficência Portuguesa de São Paulo).

Ethical Considerations

Ethics approval (CAAE: 58070622.9.0000.5483) for this study was provided by the institutional review board of the BP Hospital (Beneficência Portuguesa de São Paulo) on January 31, 2023. All eligible patients signed an informed consent form for this research protocol.

The data was collected on the RedCap platform, which has access control to ensure the security of the data collected in the research. In addition, all data was anonymized to ensure the privacy of the participants.

Sample Size

The Binomial test was used to calculate the sample size for both parts of the study. For the sampling of asymptomatic individuals (phase 1), an 80% test power was stipulated, a 70% acceptance rate of the cabin under the null hypothesis, and a 90% expected in the sampling rate were specified, resulting in 41 asymptomatic patients.

For the sampling of symptomatic individuals (phase 2), we opted for a conservative scenario. Thus, we set a power of 90%, an acceptance rate of 80% under a null hypothesis, and a 90% expected proportion in the sampling, resulting in 169 patients. Anticipating the possibility of a 12% sample loss, the researchers decided to increase the sampling to 190 symptomatic patients.

Inclusion and Exclusion Criteria

Phase 1: Asymptomatic Patients

The study's first phase involved 40 apparently healthy individuals with no clinical complaints or decompensated chronic diseases (exclusion criteria for the first phase) to assess the safety and usability of the cabin.

During this phase, all hospital staff received an email with information about the study, inviting asymptomatic individuals to participate. Interested individuals could schedule an appointment by phone to use the cabin.

Phase 2: Symptomatic Patients

The second phase included 189 individuals who could be with clinical complaints but did not meet the exclusion criteria, which were: inability to understand and answer the screening questionnaire; patients with signs and clinical symptoms indicating the need for urgent medical care, such as cardiac arrest, shock of any origin, reduced level of consciousness, focal neurological signs, epileptic seizure, chest pain, deep wound, and heavy bleeding.

During this phase, individuals who sought medical care at the institution's employee health center and who met the inclusion criteria were invited to participate. Those who agreed to participate were enrolled in the study.

Telemedicine Cabin

The telemedicine cabin Diagnostica, manufactured in Argentina, was used. It is made of plastic and fiberglass and measures 250×150×230cm dimensions. It has a forced air ventilation and filtering system, environmental lighting with intensity and color control, and an audio and video system for videoconferences (Figure 1).

Figure 1. Telemedicine cabin at the location of the research project.

The participant used his cell phone or the device available in the cabin to receive care and to interact with the person through the app developed for Android and IOS. The patient was accommodated in the cabin and selected the self-examination devices according to their needs.

The following devices were available in the cabin: an otoscope, a digital dermatoscope, and a high-definition camera, all manufactured by Firefly Global in 2021; a stethoscope Riester manufactured in 2021; and an oxyhemoglobin saturation monitor (oximeter), an electrocardiography, a heart rate and temperature monitor, and a sphygmomanometer, models PM6100, all manufactured by Berry, 2021. The patient was instructed to use the app on his cell phone, and the light signals in the cabin were added to a video display to operate the devices for the self-examination.

The cabin recorded the measurements taken by the available devices. The patient had the option, at their discretion, to join the medical consultation at any time during the use of the cabin, even if he did not use the devices. During the synchronous service, the doctor could request that the patient use any of the medical devices to obtain additional clinical information.

Doctors were recruited through direct invitations sent to professionals associated with the hospital. Selection criteria included previous experience with similar technologies and availability to participate in all stages of the study.

Data between the patient and the doctor was transmitted via the Doc24 telehealth platform, implanted in the data processing centers, and connected to the internal service network.

Study Procedure

Before starting the service, a nursing technician applied the informed consent form. After the participant's agreement, a screening questionnaire focusing on the inclusion and exclusion criteria and a questionnaire to collect demographic and socioeconomic data before using the cabin were administered. Individuals who met the exclusion criteria or had acute alterations requiring face-to-face assessment were sent to the Employee Health Center for immediate medical care.

The cabin was equipped with a Wi-Fi network and the patient should download an app to access the cabin and the interface (application) which was customized for the project.

During the appointment, the patient, seated in the cabin, can choose between selecting devices for self-examination or opting for a consultation with a doctor. At the patient's discretion, they have the option to join the medical consultation at any time during the use of the cabin, even without having used the devices.

If the participant chooses to attend a medical consultation during the service, the doctor may ask the patient to use some of the equipment to collect clinical information, or to reuse some of the equipment due to observing poor quality results initially measured by the patient.

Equipment-assisted measurements were available for medical assessment, if required, or later, in the outpatient treatment.

At the end of the service, the research participant was invited to answer the usability questionnaires translated and validated including the SUS, which is a method to measure the usability of several products and services through 10 questions answered on a scale of 1 to 5, where 1 is total disagreement and 5 is total agreement [9], and the NPS, which aims to analyze the patient's experience and satisfaction [10].

After assigning the NPS score, participants were asked to explain their ratings. Inductive thematic analysis was used to identify a rating [11]. Two authors carefully read each response individually and grouped the responses. Any disagreement was discussed with the first author, and then the response was labeled into 6 categories: "service," "usability," "difficulty of use," "innovation," "technology resistance," and "unspecific." The final step was to group the responses into potential themes.

In order to check the usability of health professionals, the participating doctors also completed the same questionnaires at the end of the daily sessions.

Statistical Analysis

Study data were collected and managed by an electronic data capture tool, REDCap (Research Electronic Data Capture), and hosted on the BP server [12,13]. REDCap is a secure software web-based platform designed to support data collection for research studies, providing (1) an intuitive interface for the validated data collection, (2) audit trails to track data manipulation and export procedures, (3) automated export procedures for further data downloads to standard statistical packages, and (4) procedures for data integration and interoperability with external sources. Once collected, data were described considering the mean and SD for the numerical variables and the absolute and relative frequencies to the categorical variables. SUS scores were interpreted using the Sauro and Lewis Curved Grading Scale [14].

Patients were classified into 2 groups according to the SUS scale score "Acceptable usability" (SUS ≥68) and "Usability issues" (SUS <68), and the percentage of patients was compared with 70% (for the Symptomatic sample), and 80% (for the Symptomatic sample) by the Binomial test. The cut-off point of 68 was selected because, despite variability in acceptability, this is the threshold at which the technology is considered to have acceptable usability.

The chi-square test or Fisher exact test was used to compare the groups' qualitative characteristics, and the Mann-Whitney Test was used for the numeric characteristics. The significance level of 0.05 was used throughout the study, and the analyses were performed using SPSS software (version 25; IBM corporation). Graphs were generated using R (version 4.3.2; R studio Team) software and the *ggplot2* package.

Results

Overview

The study was divided into 2 phases: the first occurred from March 6, 2023, to March 16, 2023, and 40 asymptomatic patients were included. The second occurred from March 17, 2023, to June 1, 2023, and 189 symptomatic patients were included.

Assessment of Asymptomatic Patients

The demographic distribution of the asymptomatic patients is shown in Table 1. Most were female, White, married, university-educated, and of higher socioeconomic status.

Concerning the use of the cabin, most of the patients requested support to use the cabin at some point during use. Only 3 patients described the reasons for requesting assistance as difficulties in using the equipment. The appointment had a duration between 10 and 52 minutes.

Upon assessing the satisfaction (Figure 2), we observed that the SUS had a mean value corresponding to Grading A+ (Table 2). When defining the cut-off point for SUS, we identified that only 5 (12.5%) patients had a score of <68 points, whereas 35 (87.5%) had a score of ≥68 points. Thus, the acceptability of 87.5% was statistically different from 70% (P=.009) and had a 95% CI 75.5%-100%, justifying the acceptability of the asymptomatic sample and allowing the study to proceed with the sampling of symptomatic patients.

Table . Distribution of demographic variables for the sampling of asymptomatic patients.

	Values (N=40)
Sex, n (%)	
Male	15 (38)
Female	25 (63)
Ethnicity and race, n (%)	
Asian	2 (5)
White	24 (60)
Mixed	8 (20)
Black	6 (15)
Education, n (%)	
High school	6 (15)
College	34 (85)
Marital status, n (%)	
Single	15 (38)
Married, common-law marriage, or partnered	24 (60)
Judicially separated or divorced	1 (3)
Monthly household income, n (%)	
Up to 3 minimum wages	8 (20)
4 - 6 minimum wages	6 (15)
7 - 11 minimum wages	13 (33)
Above 11 minimum wages	13 (33)
Care function, n (%)	
No	36 (90)
Yes	4 (10)
Has the patient requested support to use the cabin outside the previously set times?, n (%)	
No	31 (78)
Yes	9 (23)
At what point did the patient request medical advice?, n (%)	
Medical advice requested after the tests	24 (60)
Medical advice requested at the beginning; tests performed upon medical orientation	7 (18)
Did not request medical orientation	9 (23)
Age (years), mean (SD)	37.9 (10.9)
Appointment duration (minutes), mean (SD)	27.0 (10.5)
NPS[a], mean (SD)	9.6 (0.7)
SUS[b], mean (SD)	84.9 (13.0)

[a]NPS: Net Promoter Score.

[b]SUS: System Usability Scale.

Figure 2. Distribution for asymptomatic patients for the variables: (A) “How likely is it that you would recommend it to a friend or colleague.” (B) Final SUS Score upon graduation using the Curved Grading Scale. NPS: Net Promoter Score; SUS: System Usability Scale.

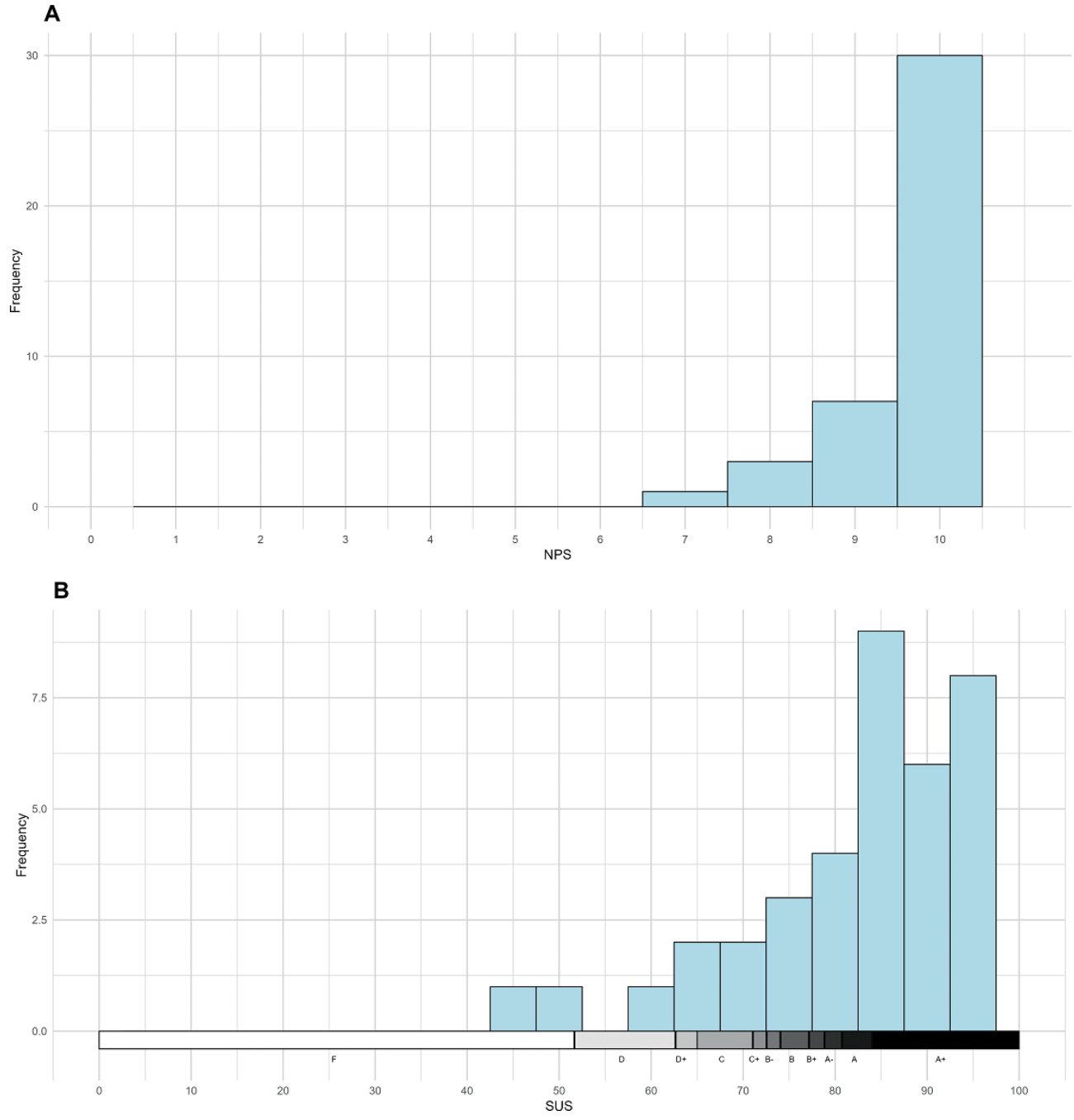


Table . Distribution of patients by Curved Grading Scale grades of asymptomatic and symptomatic groups.

SUS score range	Grading	Asymptomatic (n=40), n (%)	Symptomatic (n=189), n (%)
84.1 - 100	A+	26 (65)	128 (68.1)
80.8 - 84.0	A	3 (7.5)	8 (4.3)
78.9 - 80.7	A−	1 (2.5)	3 (1.6)
77.2 - 78.8	B+	1 (2.5)	11 (5.9)
74.1 - 77.1	B	2 (5)	8 (4.3)
72.6 - 74.0	B−	0 (0)	0 (0)
71.1 - 72.5	C+	1 (2.5)	2 (1.1)
65.0 - 71.0	C	3 (7.5)	11 (5.9)
62.7 - 64.9	C−	0 (0)	0 (0)
51.7 - 62.6	D	2 (5)	6 (3.2)
0.0 - 51.6	F	1 (2.5)	11 (5.9)

Assessment of Symptomatic Patients

The demographic distribution of the symptomatic patients is shown in Table 2. Most were female, white, single, had a college degree, had a household income up to 3 minimum wages, did not perform care functions, and had a mean age was 35.1 years.

Only 5 (2.7%) patients requested support to use the cabin outside the previously established times. The 3 reasons described for requesting support were inconsistencies in performing the cabin and difficulties in using the instruments (otoscope and oximeter) available in the cabin.

The most of participants requested medical orientation and performed the measurement under medical orientation. The duration of the appointment ranged from 2 to 58 minutes. Regarding satisfaction (Figure 3), the variable SUS mean corresponds to Grading A+ (Table 2). When assessing the participants' score in relation to the cut-off point of the SUS scale, we found that only 21 (11.1%) patients had scored <68 points, while 167 (88.8%) had scored ≥68 points. Thus, 88.8% acceptability was statistically different from 80% (P=.001) and had a 95% CI 84.3%-100%.

The 21 patients with a score of 68 or less on the SUS were considered in the group "Usability issues" at the cabin. Eleven (52.4%) were female, 13 (61.9%) self-declared White, 11 (52.4%) had a college degree, 13 (61.9%) were single, 5 (45.5%) were married, 7 (63.6%) earned up to 3 minimum wages, 8 (72.7%) did not carry out welfare activities, and had a mean age of 32.64 years (SD 8.43).

Regarding the use of the cabin by these participants, none of them required assistance from the nursing technician, 8 (38.1%) participants requested medical instruction to perform the tests, 7 (33.3%) participants requested medical orientation and made the tests under medical orientation, while 6 (28.6%) participants did not request medical orientation.

The mean appointment duration was 20.38 (SD 10.29) minutes, ranging from 6 to 40 minutes. When comparing the characteristics of patients with SUS ≥68 related to SUS <68, all of them were not statistically significant (attaining a value equal to or higher than P=.05) (Table 3), meaning that we did not identify characteristics that could have influenced the difference in the SUS score.

Figure 3. Distribution for asymptomatic patients for the variables: (A) “How likely is it that you would recommend it to a friend or colleague.” (B) Final SUS Score upon graduation using the Curved Grading Scale. NPS: Net Promoter Score; SUS: System Usability Scale.

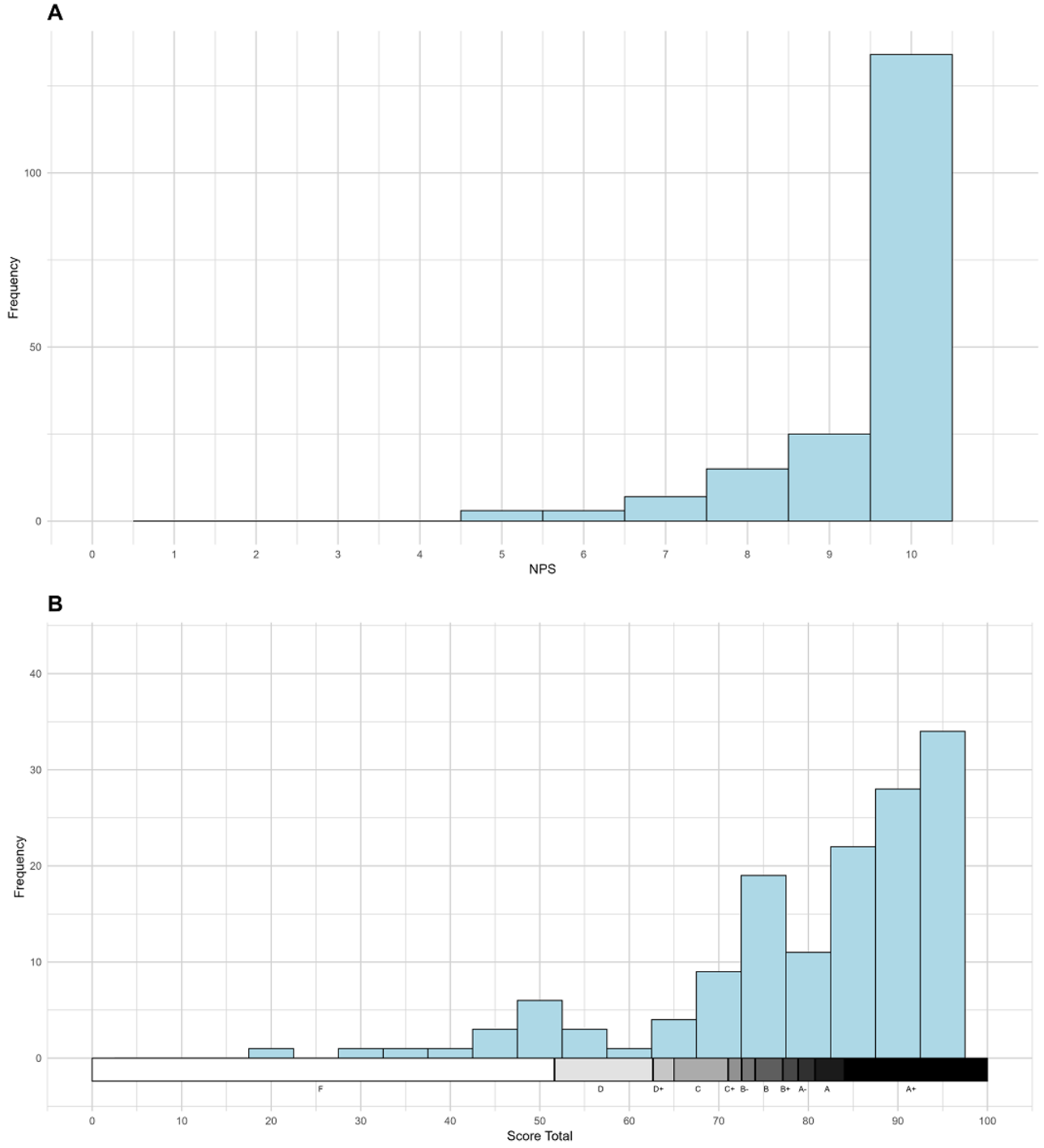


Table . Distribution of the variables for sampling symptomatic patients and comparison System Usability Scale (SUS) <68 versus ≥68.

		Patient's number	SUS[a]		*P* value
			<68	≥68	
Sex, n (%)					.05[b]
	Male	55 (29.1)	10 (47.6)	45 (26.9)	
	Female	134 (70.9)	11 (52.4)	122 (73.1)	
Ethnicity and race, n (%)					.99[c]
	Asian	3 (1.6)	0 (0)	3 (1.8)	
	White	113 (59.8)	13 (61.9)	100 (59.9)	
	Mixed	44 (23.3)	5 (23.8)	38 (22.8)	
	Black	29 (15.3)	3 (14.3)	26 (15.6)	
Education, n (%)					.62[c]
	Elementary school	6 (3.2)	1 (4.8)	5 (3)	
	High school	91 (48.1)	9 (42.9)	81 (48.5)	
	College	92 (48.7)	11 (52.4)	81 (48.5)	
Marital status, n (%)					.32[c]
	Single	93 (49.2)	13 (61.9)	80 (47.9)	
	Married, common-law marriage, or partnered	84 (44.4)	6 (28.6)	77 (46.1)	
	Widowed	2 (1.1)	0 (0)	2 (1.2)	
	Judicially separated or divorced	10 (5.3)	2 (9.5)	8 (4.8)	
Monthly household income, n (%)					.88[c]
	Up to 3 minimum wages	95 (50.3)	12 (57.1)	82 (49.1)	
	4 - 6 minimum wages	59 (31.2)	5 (23.8)	54 (32.3)	
	7 - 11 minimum wages	23 (12.2)	3 (14.3)	20 (12)	
	Above 11 minimum wages	12 (6.3)	1 (4.8)	11 (6.6)	
Care function, n (%)					.59[c]
	No	145 (76.7)	15 (71.4)	129 (77.2)	
	Yes	44 (23.3)	6 (28.6)	38 (22.8)	
Has the patient requested support to use the cabin outside the previously set times[a]?, n (%)					.99[c]
	No	184 (97.3)	21 (100)	161 (97)	
	Yes	5 (2.7)	0 (0)	5 (3)	
At what point did the patient request medical advice?, n (%)					.39[b]
	Medical advice requested after the tests	68 (36)	8 (38.1)	59 (35.3)	
	Medical advice requested at the beginning; tests performed upon medical orientation	85 (45)	7 (33.3)	78 (46.7)	
	Did not request medical orientation	36 (19)	6 (28.6)	30 (18)	
Age (years), mean (SD)		35.1 (10.9)	35.0 (10.2)	35.2 (11.0)	.93
Appointment duration (minutes), mean (SD)		21.8 (11.8)	20.4 (10.3)	22.0 (12.0)	.58

	Patient's number	SUS[a]		*P* value
		<68	≥68	
NPS[d], mean (SD)	9.4 (1.3)	7.7 (2.5)	9.6 (0.8)	<.01
SUS, mean (SD)	86.0 (15.5)	51.2 (11.5)	90.3 (9.1)	<.01

[a]There was 1 missing value for the variable "SUS" and 1 missing value for the variable "Has the patient requested support to use the cabin outside the previously set times?".

[b]Chi-square test.

[c]Fisher exact test.

[d]NPS: Net Promoter Score.

Regarding the reasons for their ratings, only 3 participants chose not to answer this question. After careful reading of each response by 2 researchers, the answers were grouped according to the following categories:

- Service: assessment of the interaction between the cabin's professionals and the patient, considering the attention to the needs, clarification of doubts, and resolution of problems. Examples: (1) "Doctor's attention and understanding of my problem"; (2) "The speed and attention of the professionals who cared for me"; and (3) "Spectacular service. All my questions were answered. I thought it was excellent."
- Usability: an assessment of the ease of use of both the interface and the equipment. Examples: (1) "The experience was excellent; I enjoyed doing the tests myself and talking to the doctor about them"; (2) "Super practical, efficient, easy to use the equipment to diagnose, and talk to the doctor afterwards. Loved it!"; and (3) "I found it practical, accessible, and easy to use."
- Difficulty of use: difficulty in using equipment and tools to locate and understand the functions. Examples: (1) "Difficulty in accurately showing images and pain sites to the professional and the client's lack of affinity with the equipment can make it difficult to reach a diagnosis"; (2) "Some people may have difficulty with technology"; and (3) "Conflicting instructions between the cell phone and the screen at the self-diagnosis."
- Innovation: considerations about new products and technologies that can be aggregated to a business or project to bring about improvements. Examples: (1) "Great technology"; (2) "Very advanced technology"; and (3) "For being an innovative, versatile, and practical project."
- Technology Resistance: resistance to technological change. Example: "In no way does it replace the personal contact with the doctor, as there is no way of knowing for sure what you are feeling."
- Nonspecific: answers that were not intended or that did not belong to a predetermined group or situation, in this case, to classes of sentences mentioned above. Examples: (1) "Ok"; (2) "I loved it"; and (3) "Top."

The distributions of the classifications are shown in Table 4.

Most opinions are classified as related to "Usability." When comparing the classes of sentences of those who scored SUS equal or less than 68 (usability issues) with the remaining (acceptable usability), we found statistical differences in the class "Difficult Usability." In the group "Usability issues," 5 (27.8%) scored in this class, while 8 (4.8%) scored in the group "Acceptable usability."

Table . Distribution of the classes of sentences.

	Patient's number, n (%)	System Usability Scale (SUS), n (%)		*P* value
		<68	≥68	
Service				.06[a]
No	106 (57)	14 (77.8)	92 (55.1)	
Yes	80 (43)	4 (22.2)	75 (44.9)	
Usability				.23[a]
No	89 (47.8)	11 (61.1)	77 (46.1)	
Yes	97 (52.2)	7 (38.9)	90 (53.9)	
Difficulty of use				<.01[b]
No	173 (93)	13 (72.2)	159 (95.2)	
Yes	13 (7)	5 (27.8)	8 (4.8)	
Innovation				.99[b]
No	175 (94.1)	17 (94.4)	157 (94)	
Yes	11 (5.9)	1 (5.6)	10 (6)	
Technological resistance				.10[b]
No	185 (99.5)	17 (94.4)	167 (100)	
Yes	1 (0.5)	1 (5.6)	0 (0)	
Nonspecific				.17[b]
No	157 (84.4)	13 (72.2)	143 (85.6)	
Yes	29 (15.6)	5 (27.8)	24 (14.4)	

[a]Chi-square test.

[b]Fisher exact test.

Medical Assessment

At the end of the day's work, the doctor performing the appointments was oriented to answer the NPS and SUS, evaluating his or her day's work. The assessment was performed by 2 doctors, reducing the variability of the influence of the medical service on usability; 89% of the services were performed by one of the doctors, and 11% were performed by the other. The distribution of the NPS and SUS is shown in Table 5.

The main difficulties were related to technical difficulties, such as software updates and network connection problems, besides difficulties in adjusting the cabin's technology (problems with the sensors for measuring the weight and height; problems with the image or screen in the cabin).

Figures 4 and 5 show the evolution of the allocation of medical grades over time.

Table . Distribution of the Net Promoter Score (NPS) and System Usability Scale (SUS) values answered by the doctor at the end of each day of service.

	Patient's number (N)	Mean (SD)	95% CI	Min[a]	1oQ[b]	Medium	3oQ[c]	Max[d]
NPS	41	8.0 (1.4)	7.6-8.4	5.0	7.0	8.0	9.0	10.0
SUS	41	67.8 (13.6)	63.5-72.1	45.0	57.5	65.0	72.5	95.0

[a]Min: minimum.

[b]1oQ: first quartile.

[c]3oQ: third quartile.

[d]Max: maximum.

Figure 4. Evolution of the NPS answered by the doctor at the end of each day of service. NPS: Net Promoter Score.

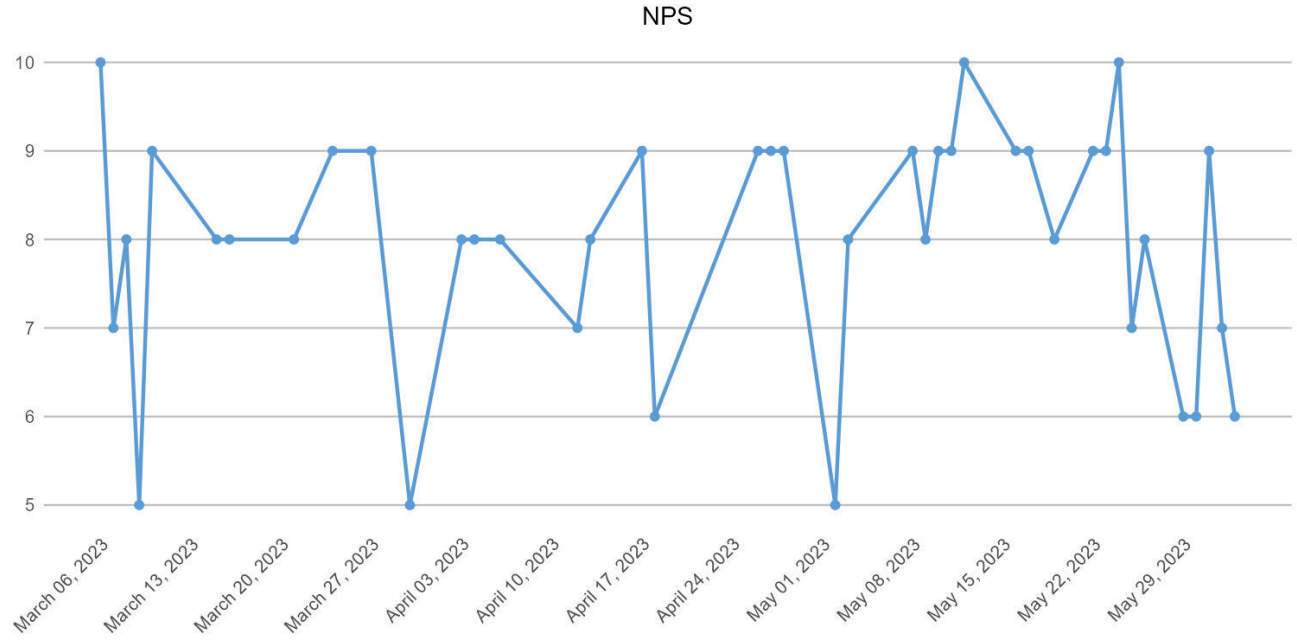
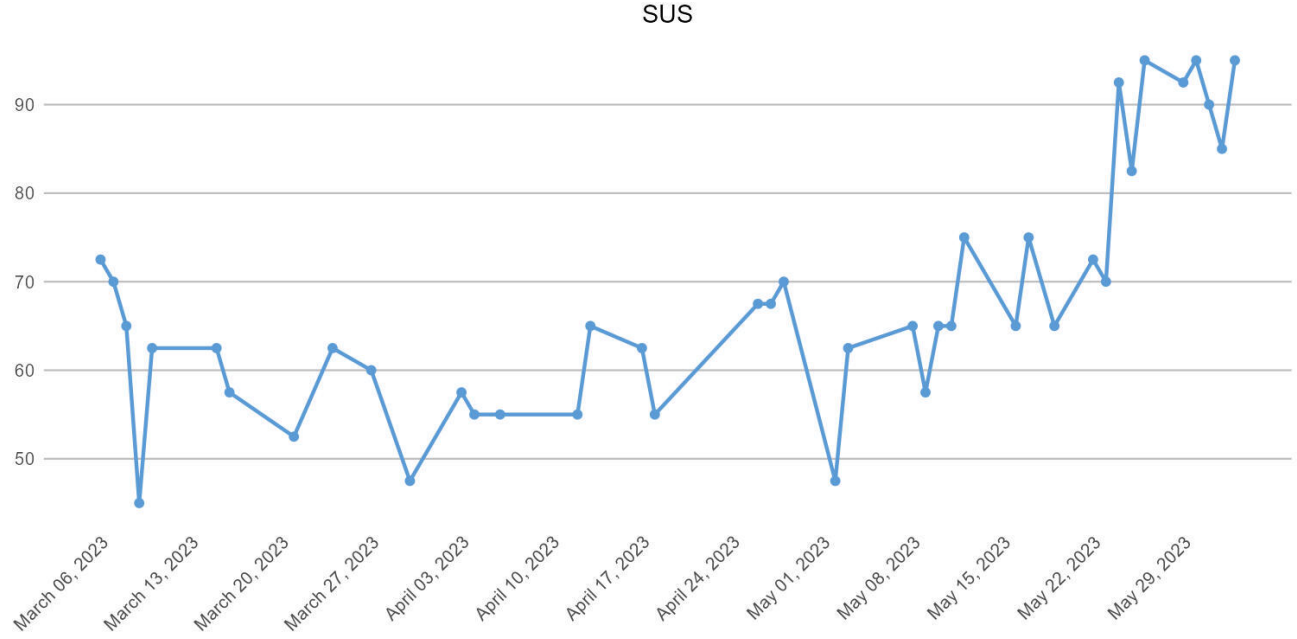


Figure 5. Evolution of the SUS score answered by the doctor at the end of each day of service. SUS: System Usability Scale.



Discussion

Principal Findings

The assessment of the implementation of the telehealth cabin in this study aimed to observe the level of favorability and usability of users (doctors and patients) to the cabin and the self-examination devices available, added to their experience of remote service through the technologies available.

The analysis of the results showed that the participants of the first phase of the research (asymptomatic patients) had a different socioeconomic profile than those of the second phase (symptomatic patients), with higher education and monthly income, probably due to the new technology, which attracted leaders and directors of the institution to the research. Although this group was only used to assess the acceptability of both the cabin and the self-examination devices in order to continue the study with symptomatic patients, the outcomes of the favorability and usability were similar between the groups.

Therefore, despite the disparity in access to technology among populations of different economic levels, such a result suggests that the telehealth cabin may have broad applicability for several publics due to the increasing availability and use of mobile devices among the whole population, thus promoting health the equity and reducing social inequalities.

A study with similar equipment performed with French students also concluded that the personal relationship with the technology did not influence the intended use of the telemedicine cabin [15].

A significant portion of the participants did not work in the hospital care sector and therefore did not routinely use the self-examination devices, suggesting that lack of familiarity with the use of the self-examination devices may not impact the use of the cabin. However, this study was performed with professionals working in a hospital, which may bring a certain level of familiarity with those devices, even for those not directly working in the care sector.

Still, we observed that more than half of the patients requested an interview with the doctor, either before or after self-examination, indicating the relevant role of the doctor in such a kind of service.

High favorability and usability were observed for both patient groups (symptomatic and asymptomatic). No statistically significant differences were found when comparing the socioeconomic and cultural profile of those patients with the group of symptomatic who considered the cabin to have acceptable usability, which shows that the usability and favorability of the cabin are unrelated to socioeconomic and cultural factors. However, concerning comments from this group show that some aspects were considered positive even in this group. Furthermore, only one patient described a comment that could be classified in the category "Resistance to Technology," suggesting that most of the population tested is interested in learning and adopting technologies that improve their routines and experiences.

The findings of this study support the work by Scott Kruse et al [8], asserting that the principal barriers to telemedicine are technology specific. They can be overcome by continuous improvement of these technologies, adequate training of technology users, and personal interaction between the patient and the care provider [8].

Concerning the favorability and usability of doctors who performed the cabin services cabin, we found that the answer was also positive, but with a slightly lower level of usability than that of the patients. This fact seems to agree with other studies that assert that some health professionals may be reluctant. However, this resistance may be related to the difficulty of the technology. Therefore, both the favorability and usability can improve with the advancement of technology and the due training of doctors in web-based physical tests [16,17]. This phenomenon highlights the importance of training programs as a strategy to improve user experience and expedite the adoption of this technology.

A limitation of this study is the participation of only individuals without respiratory symptoms (a consequence of the COVID pandemic) or without symptoms that could indicate a medical emergency. Another limiting factor is the poor accessibility of the cabin for people with disabilities. Furthermore, because the total sampling included people employed at the hospital, and so, in the productive age group and predominantly female [18], along with workers in the day shift, this study may have introduced a bias in the perception of favorability and usability of the participants (for instance, older men are less prone to participate in several telehealth activities [19]).

Finally, we emphasize that the results obtained from health professionals should also be interpreted with caution due to the limited number of professionals who provided care for the study.

It should also be noted that no studies using this type of technology were identified in South America to compare the results obtained. In addition, the fact that this study was conducted after the pandemic may have influenced the results given the impact of the COVID-19 pandemic on people's lives, which may have affected perceptions.

Based on the favorable assessment of the technology by the participants in this study (patients and doctors) new studies that extend its use to different populations, without any link to the hospital and from different sectors of activity, different work shifts, sex, and different age groups, may bring new inputs to the application of such technology, expanding the service types and favoring the inclusion of the service in remote locations and for diverse populations.

Conclusions

The authors conclude that the telehealth cabin had good usability and favorability by patients, regardless of the socioeconomic and cultural profile of the population. Doctors who performed the cabin services attested to the cabin's good usability cabin but with a usability index slightly lower than patients. Most patients used the devices for self-examination, with little need for technical assistance. However, a SUS score greater than 68 does not necessarily mean that there are no usability issues to be addressed.

The outcomes of this research allow us to assess as feasible the expansion of this kind of practice from the point of view of usability and that this study can contribute as a subsidy to the construction of public policies to expand the access of the Brazilian population to qualified medical care, promoting the engagement of patients in their health, providing health professionals with an extended view of the clinical conditions through embedded devices even in areas with a shortage of this type of assistance.

Acknowledgments

All authors would like to thank the Votorantim Institute for subsidizing the possibility of this project and for partnering with DOC24 and Diagnóstica, which brought a new way of experiencing the health of the future through innovative equipment.

Conflicts of Interest

None declared.

References

1. Universal Declaration of Human Rights [Web page in Portugese]. UNICEF. 1948. URL: https://www.unicef.org/brazil/declaracao-universal-dos-direitos-humanos [accessed 2022-02-20]
2. Jacobs B, Ir P, Bigdeli M, Annear PL, Van Damme W. Addressing access barriers to health services: an analytical framework for selecting appropriate interventions in low-income Asian countries. Health Policy Plan 2012 Jul;27(4):288-300. [doi: 10.1093/heapol/czr038] [Medline: 21565939]
3. Viacava F, Laguardia J, Ugá MAD, Porto S. PROADESS - performance assessment of the Brazilian health system: indicators for monitoring (research report). PROADESS. URL: http://www.proadess.icict.fiocruz.br/Relatorio_Proadess_08-10-2012.pdf [accessed 2023-07-18]
4. Silva AB, da Silva RM, Ribeiro GDR, et al. Three decades of telemedicine in Brazil: mapping the regulatory framework from 1990 to 2018. PLoS ONE 2020;15(11):e0242869. [doi: 10.1371/journal.pone.0242869] [Medline: 33237947]
5. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010 Jun 16;12(2):e21. [doi: 10.2196/jmir.1357] [Medline: 20554500]
6. Polisena J, Coyle D, Coyle K, McGill S. Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations. Int J Technol Assess Health Care 2009 Jul;25(3):339-349. [doi: 10.1017/S0266462309990201] [Medline: 19619353]
7. Russo JE, McCool RR, Davies L. VA telemedicine: an analysis of cost and time savings. Telemed J E Health 2016 Mar;22(3):209-215. [doi: 10.1089/tmj.2015.0055] [Medline: 26305666]
8. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12. [doi: 10.1177/1357633X16674087] [Medline: 29320966]
9. Brooke J. SUS: a quick and dirty usability scale. In: Usability Evaluation In Industry: CRC Press; 1995:189-194 URL: http://hell.meiert.org/core/pdf/sus.pdf [accessed 2025-04-03]
10. Reichheld FF. The one number you need to grow. Harv Bus Rev 2003 Dec;81(12):46-54. [Medline: 14712543]
11. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
12. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381. [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
13. Harris PA, Taylor R, Minor BL, et al. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95(103208):103208. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
14. Sauro J, Lewis JR. Quantifying the User Experience: Practical Statistics for User Research: Morgan Kaufmann; 2016.
15. Baudier P, Kondrateva G, Ammi C. The future of telemedicine cabin? The case of the French students' acceptability. Futures 2020 Sep;122(102595):102595. [doi: 10.1016/j.futures.2020.102595] [Medline: 32834075]
16. Lisboa KO, Hajjar AC, Sarmento IP, Sarmento RP, Gonçalves SHR. The history of telemedicine in Brazil: challenges and advantages. Saude Soc 2023;32(1):e210170pt. [doi: 10.1590/s0104-12902022210170pt]
17. Ladin K, Porteny T, Perugini JM, et al. Perceptions of telehealth vs in-person visits among older adults with advanced kidney disease, care partners, and clinicians. JAMA Netw Open 2021 Dec 1;4(12):e2137193. [doi: 10.1001/jamanetworkopen.2021.37193] [Medline: 34870680]
18. Machado MH, Wermelinger M, Machado AV, et al. Profile and working conditions of health professionals in times of Covid-19: the Brazilian reality. In: Portela MC, Reis LGC, Lima SML, editors. Covid-19: Challenges for the Organization and Repercussions on Health Systems and Services: Fiocruz Publishing House; 2022:283-295.
19. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the health information national trends survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172. [doi: 10.2196/jmir.3117] [Medline: 25048379]

Abbreviations

BP Hospital: Beneficência Portuguesa de São Paulo
NPS: Net Promoter Score
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by A Kushniruk; submitted 21.12.23; peer-reviewed by C Udeh, J Brooke, NS Sood; revised version received 02.01.25; accepted 05.01.25; published 16.04.25.

Please cite as:
de França CVG, Segalla PB, Reis FSDA, Pereira JRS, de Mattos AO, Ferron RDM, de Oliveira CZ, Borges JB, Hoffmann LQ, Caboclo EDG
Patients' and Physicians' Experience With and Acceptability of a Telemedicine Cabin: Mixed Methods Study
JMIR Hum Factors 2025;12:e55430
URL: https://humanfactors.jmir.org/2025/1/e55430
doi:10.2196/55430

Gender-Inclusive Language in Public-Facing Labor and Delivery Web Pages in the New York Tristate Area: Cross-Sectional Study

Sarah Mohsen Isaac[1], BA; Mark Dawes[2], BSc; Emily Ruth Howell[1,3], DO; Antonia Francis Oladipo[1,4], MSCI, MD

[1]Hackensack Meridian School of Medicine, 123 Metro Blvd, Nutley, NJ, United States

[2]Cornell University, Ithaca, NY, United States

[3]Department of Obstetrics and Gynecology, Hackensack University Medical Center, Hackensack, NJ, United States

[4]Division on Maternal-Fetal Medicine, Department of Obstetrics and Gynecology, Hackensack University Medical Center, Hackensack, NJ, United States

Corresponding Author:
Sarah Mohsen Isaac, BA
Hackensack Meridian School of Medicine, 123 Metro Blvd, Nutley, NJ, United States

Abstract

Background: Transgender and nonbinary (TGNB) individuals are increasingly intentionally becoming pregnant to raise children, and hospital websites should reflect these trends. For prospective TGNB parents, a hospital website is the only way they can assess their safety from discrimination while receiving perinatal care. Cisnormativity enforced by communication gaps between medical institutions and TGNB patients can and has caused delays in receiving urgent care during their pregnancy.

Objective: The aim of this study was to evaluate the current prevalence of gender-inclusive terminology among labor and delivery services in the New York tristate area.

Methods: The labor and delivery web pages of 189 hospitals from New York, New Jersey, and Connecticut were examined for gender-inclusive language. “Fully inclusive” websites explicitly acknowledged lesbian, gay, bisexual, transgender, queer, intersex, and asexual plus other gender- and sexual-oriented (LGBTQIA+) parents, “inclusive” websites did not use gendered terminology for parents, and “noninclusive” websites used gendered terms at least once in the text reviewed. The hospitals’ web pages were further stratified by Healthcare Equality Index scores and population classifications defined by the 2013 National Center for Health Statistics Urban-Rural classification given to the county that each hospital was located in.

Results: Of the 300 hospital websites reviewed, only 189 websites met the criteria for inclusion. Overall, only 6.3% (n=12) of labor and delivery web pages were “inclusive” or “fully inclusive.” No geographic areas (*P*=.61) or Healthcare Equality Index scores (*P*=.81) were associated with inclusive or fully inclusive language.

Conclusions: Hospitals need to use inclusive language to help TGNB people identify hospitals where their existence and needs are acknowledged and thus feel more comfortable in their transition to parenthood.

(JMIR Hum Factors 2025;12:e53057) doi:10.2196/53057

KEYWORDS

OBGYN; transgender; nonbinary; pregnancy; maternity; transmasculine; observational study; gestational; perinatal care; communication; labor; USA; United States; New York City; sexual orientation; inclusion; parents; obstetrician gynecologist; delivery

Introduction

Within the last 2 decades, as transgender and nonbinary (TGNB) people have gained greater legal and social recognition, those in the TGNB community who want to become pregnant have become increasingly more common, as have the number of families with same-sex, transgender, or nonbinary parents [1]. Despite this trend, the field of medicine has mostly maintained the heteronormative model of a “mother” and “father” as opposed to a more fluid, freeform reality that accompanies the rise of lesbian, gay, bisexual, transgender, queer, intersex, and asexual plus other gender- and sexual-oriented (LGBTQIA+) couples who raise children.

As pregnant TGNB people attend numerous prenatal visits, they make repeated contact with heterosexist health care systems without the ability to hide their transgender status. Pharr [2] explains that heterosexism is not an active form of discrimination but rather “a belief that the world is and must be heterosexual.” According to the heterosexist worldview, every couple contains—or should contain—only 2 gender-conforming people of the opposite sex [3].

Heterosexism works with homophobia to make health care inaccessible for LGBTQIA+ populations. Current literature has

found that these parents and couples are often invalidated and marginalized throughout the health care process through obstacles like registration forms, comments from ancillary staff, and physicians who are unprepared to deal with LGBTQIA+ couples [4-6]. These experiences can take a deep mental and emotional toll—doubly so during the sensitive transition to parenthood [7].

Beyond these mental health impacts, the invisibility of TGNB parents can negatively affect pregnancy outcomes. LGBTQIA+ patients were less likely to trust providers and divulge important medical information when they received heteronormative medical treatment [8,9]. Patients may also experience delays receiving urgent pregnancy care due to the systematic heterosexism built into the health care system.

For example, Parker et al [6] discuss the experiences of a transmasculine patient who was delayed in seeing the doctor because the receptionist argued that he was not the intended person. These patterns are not limited to individual heterosexist health care workers but also integrated into hospital software. Berger et al [10] describe another scenario where a transmasculine patient's care was delayed, this time because the hospital's electronic medical record required that he reregister as female to document the pregnancy, regardless of his actual gender identity. A delay in care to circumvent cisnormative systems can be dangerous for all pregnant TGNB people. Ultimately, discrimination has often necessitated that prospective TGNB parents discern a "safe" hospital before seeking care.

Historically, LGBTQIA+ people have relied on word of mouth from their personal social circles to find safer health care [11]. However, younger LGBTQIA+ people, especially ones without LGBTQIA+ support networks, also rely on the internet to search for health information and providers that are inclusive of LGBTQIA+ people [12,13]. This vetting of hospitals, along with the increase of patient choice and consumerism for perinatal care in general, has prompted hospitals to advertise their unique benefits, such as low cesarean rates, "baby-friendly" designations, and private rooms [14-19]. Hospitals have tailored their advertisements for other demographics around them, but there is a dearth of literature showing how hospitals advertise their services for LGBTQIA+ populations, who rely on publicly available information to find inclusive care and preserve their health and safety [20].

Purdie-Vaughns et al [21] point to purposeful word choice as one safety cue that, when recognized, signals protection from identity-based discrimination. Hospitals might therefore attract pregnant LGBTQIA+ parents by crafting more inclusive obstetrical web pages. These pages could signal inclusivity by explicitly referencing LGBTQIA+ care or by avoiding gender-exclusive language like "mother and baby," "mom," or presumptive she/her pronouns for parents. Through the words chosen on these public-facing web pages, hospitals thus enable parents to choose to give birth in places where their existence is actively supported during the physically dangerous and psychologically difficult transition to parenthood.

The states surrounding New York City—New York, New Jersey, and Connecticut—boast an exceptionally high population density of LGBTQIA+ individuals, who make up between 3% to 5% of the total adult population [22]. This geographical region is viewed as more inclusive towards LGBTQIA+ people than average, so hospitals may have more incentive to provide inclusive care [23,24]. This study aims to evaluate the current prevalence of LGBTQIA+ inclusion and gender-inclusive terminology among labor and delivery (L&D) service web pages in the New York tristate area.

Methods

Study Design

The targeted words used to assess gender-inclusiveness for this study were largely adapted from the Jennings et al [25] study of gender-inclusive language on National Health Service websites.

The official public-facing obstetric web pages of nonfederal, short-term, acute-care hospitals from Connecticut, New Jersey, and New York were analyzed (n=300). Hospitals without L&D services or web content describing these services were excluded (n=189). Websites were reviewed from late November 2022 to January 2023.

Hospitals were categorized by state, 2013 National Center for Health Statistics Rural-Urban classification, and Healthcare Equality Index (HEI) score. The National Center for Health Statistics Rural-Urban classification is a tool used to identify urban and rural areas of the United States. It was used to analyze any association between urbanization and hospital-based inclusiveness of LGBTQIA+ people. The HEI score is the national LGBTQIA+ benchmarking criterion that assesses health care facilities' policies and practices regarding equity and inclusion of LGBTQIA+ patients, visitors, and employees. It was also used to identify if there was any association between a hospital's publicly perceived LGBTQIA+ inclusivity and word choice on the web pages.

For each hospital, at least 1 web page was examined alongside up to 2 additional pages as supplementation for language analysis. The gendered language used was recorded and analyzed by a single reviewer. The complete L&D-related text was analyzed and the types of gendered language used were recorded. (Explicit discussion of related services, such as chestfeeding, within the same site was excluded.) Any non–gender-inclusive descriptors for the name of the building or third-party services were also excluded from analysis, as these are often not controlled by hospital administration.

Language Analysis

Each web page was reviewed independently by the chief reviewer to minimize any discrepancies. Each hospital's L&D web page was rated as "fully inclusive," "inclusive," or "noninclusive." "Fully inclusive" websites explicitly acknowledged LGBTQIA+ or TGNB parents. "Inclusive" websites did not use gendered terminology or pronouns for prospective parents. "Noninclusive" websites used the terms "woman" or "women"; "mom" or "mother"; other terms for women; "father" or "dad"; or she/her pronouns at least once in the text reviewed.

Statistical Analysis

Categories were analyzed using χ^2 tests presented as frequencies with percentages. *P* values <.05 were considered statistically significant, and all tests were 2-sided.

Ethical Considerations

Ethics and insitutional review board approval were not required since the study did not include human or animal subjects and all data were collected from publicly available websites.

Results

Of the 300 hospital websites reviewed, 111 hospital websites did not have a L&D web page or did not have content describing their L&D services (Multimedia Appendix 1). Of the remaining 189 websites analyzed, 12 (6.3%) of them used fully inclusive or inclusive language (Table 1). Only 1 hospital (0.5%) was considered fully inclusive because it acknowledged “same-sex” couples in its L&D content. The most common noninclusive terms used were “mom” or “mother” (n=166, 87.8%) and “woman” or “women” (n=94, 49.7%). No geographic areas (*P*=.61) or HEI scores (*P*=.81) were associated with inclusive or fully inclusive language (Tables 2-4).

Table . Labor and delivery web pages that used each type of language (N=189).

	Fully inclusive, n (%)	Inclusive, n (%)	Noninclusive, n (%)	Total, n (%)
Websites	1 (0.5)	11 (5.8)	177 (93.7)	189 (100)
Language used				
woman or women	1 (1.1)	0 (0)	93 (49.2)	94 (49.7)
mother or mom	0 (0)	0 (0)	166 (87.8)	166 (87.8)
she/her (parent)	0 (0)	0 (0)	42 (22.2)	42 (22.2)
she/her (staff)	0 (0)	0 (0)	9 (4.8)	9 (4.8)
synonyms for women (ladies, etc)	0 (0)	0 (0)	1 (0.5)	1 (0.5)
father or dad	0 (0)	0 (0)	33 (17.5)	33 (17.5)

Table . Summary of HEI[a] scores and gender-inclusive language used on L&D[b] web pages .

	Fully inclusive (n=1), n (%)	Inclusive (n=11), n (%)	Noninclusive (n=177), n (%)	Total (N=189), n (%)
HEI score=100%	0 (0)	5 (45.5)	60 (33.9)	65 (34.8)
HEI score <100%	0 (0)	1 (9.1)	25 (14.1)	26 (13.9)
HEI score not applicable	1 (100)	5 (45.5)	92 (52)	98 (52.4)

[a]HEI: Healthcare Equality Index.

[b]L&D: labor and delivery.

Table . Summary of gender-inclusive language used on labor and delivery web pages and population data.

	Fully inclusive (n=1), n (%)	Inclusive (n=11), n (%)	Noninclusive (n=177), n (%)	Total (N=189), n (%)
Large central metro	0 (0)	6 (54.5)	53 (30.6)	59 (31.2)
Large fringe metro	0 (0)	3 (27.3)	65 (37.6)	68 (36)
Medium metro	1 (100)	1 (9.1)	31 (17.9)	33 (17.5)
Small metro	0 (0)	0 (0)	7 (4)	7 (3.7)
Micropolitan	0 (0)	1 (9.1)	17 (9.8)	18 (9.5)
Noncore	0 (0)	0 (0)	4 (3.9)	4 (2.1)

Table . Examples of suggested gender-inclusive language [19,25].

Non–gender-inclusive language	Gender-inclusive language
“Mothers”	“Birthing parents” OR “women and birthing parents”
“Pregnant woman”	“Pregnant patient” OR “pregnant person”
“Mother and baby unit”	“Maternity unit” OR “birthing unit”

Discussion

Principal Findings

These results demonstrate that there is a large barrier for TGNB parents to search for and identify potentially inclusive pregnancy care. Out of the 12 inclusive and fully inclusive L&D web pages, 92% were inclusive not because they included gender-additive language or LGBTQIA+ topics but rather because they omitted the pregnant person's gender altogether by addressing the reader in the second person. The websites may have been inclusive not by intention but by coincidence. In stark contrast, there are multiple private reproductive endocrinology and infertility clinics that specifically target LGBTQIA+ couples using specific gender-inclusive language on their websites [26]. TGNB parents who are accustomed to a purposefully inclusive experience during their fertility journey and early pregnancy may be caught off guard by the sudden invisibility of their identities as they progress further through their pregnancy.

Interestingly, none of the hospitals that are acknowledged for their excellence in LGBTQIA+ care in other specialties discussed serving prospective TGNB parents for L&D care on their websites. This likely reflects a wider societal trend of "repronormativity," by which society at large does not recognize reproductive sex between TGNB parents as possible or legitimate [27].

The accessibility of websites and web-based platforms is important for TGNB people to find services and connect to similar parents. Our above findings suggest less than 10% of hospitals use gender-inclusive language when representing their services. Thus, TGNB parents who are not connected to a wider LGBTQIA+ community may struggle to find inclusive prenatal care and delivery services due to the lack of representation.

Limitations

The use of multiple surrogate end points may limit this study. The analyzed web pages, while used as a proxy for the culture in L&D departments, may not fully represent institutional attitudes and practices once parents start using their providers. This is exacerbated by delays between institutional attitude changes and hospital website updates. HEI scores were ineffective in predicting the LGBTQIA+ inclusivity of hospitals' web pages because HEI scores are determined purely through institutional measures like nondiscrimination policies; they do not directly address subtler, underlying heterosexism that hopeful TGNB parents try to avoid in their health care. Finally, the methodology used in this study makes it impossible to establish a causal link between gendered terminology and the quality of LGBTQIA+ inclusive care.

Additionally, it is important to note that individual TGNB people may feel varying levels of dysphoria around maternal terms; some TGNB parents may not consider the words "mom" or "mother" to be exclusively for women. However, using gender-inclusive language and terminology is an important step towards providing a more welcoming and inclusive environment for TGNB parents, regardless of those individuals' personal dysphoria triggers.

Conclusions

Using gender-inclusive language and terminology is the first step towards providing a more welcoming and inclusive environment for pregnant TGNB parents. Hospitals that want to be recognized as more inclusive towards LGBTQIA+ people can integrate gender-additive language into their L&D web pages (eg, "mothers and birthing parents") rather than omit mentions of gender entirely [28]. US hospitals should consider expanding this language to meet the needs of a growing group of people who are having children. Future research should be done including LGBTQIA+ patient advocate groups on the use of inclusive language within health care providers' obstetrical and gynecology departments, specifically in L&D units and on how this language impacts TGNB parents' health outcomes and rapport with physicians.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of hospitals reviewed for analysis, along with a list of hospitals where labor and delivery pages were missing for various reasons.
[XLSX File, 77 KB - humanfactors_v12i1e53057_app1.xlsx]

References

1. Tornello SL, Bos H. Parenting intentions among transgender individuals. LGBT Health 2017 Apr;4(2):115-120. [doi: 10.1089/lgbt.2016.0153] [Medline: 28212056]
2. Pharr S. Homophobia: A Weapon of Sexism: Chardon Press; 1988.
3. Hudak N. "Who's the Mom?": heterosexism in patient-provider interactions of queer pregnant couples. Health Commun 2023 Jan;38(1):114-123. [doi: 10.1080/10410236.2021.1936752] [Medline: 34096428]
4. Wojnar DM, Katzenmeyer A. Experiences of preconception, pregnancy, and new motherhood for lesbian nonbiological mothers. J Obstet Gynecol Neonatal Nurs 2014;43(1):50-60. [doi: 10.1111/1552-6909.12270] [Medline: 24354595]

5. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. Obstet Gynecol 2014 Dec;124(6):1120-1127. [doi: 10.1097/AOG.0000000000000540] [Medline: 25415163]
6. Parker G, Ker A, Baddock S, Kerekere E, Veale J, Miller S. "It's Total Erasure": trans and nonbinary peoples' experiences of cisnormativity within perinatal care services in aotearoa New Zealand. Womens Reprod Health (Phila) 2022 Dec 26;10(4):591-607. [doi: 10.1080/23293691.2022.2155496]
7. Greenfield M, Darwin Z. Trans and non-binary pregnancy, traumatic birth, and perinatal mental health: a scoping review. Int J Transgend Health 2021;22:203-216. [doi: 10.1080/26895269.2020.1841057] [Medline: 34806082]
8. Utamsingh PD, Richman LS, Martin JL, Lattanner MR, Chaikind JR. Heteronormativity and practitioner-patient interaction. Health Commun 2016;31(5):566-574. [doi: 10.1080/10410236.2014.979975] [Medline: 26421354]
9. Seelman KL, Colón-Diaz MJP, LeCroix RH, Xavier-Brier M, Kattari L. Transgender noninclusive healthcare and delaying care because of fear: connections to general health and mental health among transgender adults. Transgend Health 2017;2(1):17-28. [doi: 10.1089/trgh.2016.0024] [Medline: 28861545]
10. Berger AP, Potter EM, Shutters CM, Imborek KL. Pregnant transmen and barriers to high quality healthcare. Proc Obstet Gynecol 2015;5(2):1-12. [doi: 10.17077/2154-4751.1285]
11. Hudak N, Bates BR. In pursuit of "queer-friendly" healthcare: an interview study of how queer individuals select care providers. Health Commun 2019 Jul;34(8):818-824. [doi: 10.1080/10410236.2018.1437525] [Medline: 29482359]
12. Charest M, Kleinplatz PJ, Lund JI. Sexual health information disparities between heterosexual and LGBTQ+ young adults: implications for sexual health. Can J Hum Sex 2016;25(2):74-85. [doi: 10.3138/cjhs.252-A9]
13. Mitchell KJ, Ybarra ML, Korchmaros JD, Kosciw JG. Accessing sexual health information online: use, motivations and consequences for youth with different sexual orientations. Health Educ Res 2014 Feb;29(1):147-157. [doi: 10.1093/her/cyt071] [Medline: 23861481]
14. Fischer S, Pelka S, Riedl R. Understanding patients' decision-making strategies in hospital choice: literature review and a call for experimental research. Cog Psych 2015 Dec 15;2(1). [doi: 10.1080/23311908.2015.1116758]
15. Mehrotra A, Wolfberg A, Shah NT, et al. Impact of an educational program and decision tool on choice of maternity hospital: the delivery decisions randomized clinical trial. BMC Pregnancy Childbirth 2022 Oct 10;22(1):759. [doi: 10.1186/s12884-022-05087-y] [Medline: 36217115]
16. Witteveen AB, De Cock P, Huizink AC, et al. Pregnancy related anxiety and general anxious or depressed mood and the choice for birth setting: a secondary data-analysis of the DELIVER study. BMC Pregnancy Childbirth 2016 Nov 22;16(1):363. [doi: 10.1186/s12884-016-1158-7] [Medline: 27871257]
17. Lally S, Lewis V. Maternity care patient engagement strategies. California Health Care Foundation. 2014. URL: https://www.chcf.org/wp-content/uploads/2017/12/PDF-MaternityCarePtEngagementStrategiesIHA.pdf [accessed 2024-12-13]
18. Tayyari Dehbarez N, Lou S, Uldbjerg N, Møller A, Gyrd-Hansen D, Søgaard R. Pregnant women's choice of birthing hospital: a qualitative study on individuals' preferences. Women Birth 2018 Dec;31(6):e389-e394. [doi: 10.1016/j.wombi.2017.11.006] [Medline: 29198502]
19. Maurer M, Firminger K, Dardess P, Ikeler K, Sofaer S, Carman KL. Understanding consumer perceptions and awareness of hospital-based maternity care quality measures. Health Serv Res 2016 Jun;51 Suppl 2(Suppl 2):1188-1211. [doi: 10.1111/1475-6773.12472] [Medline: 26927831]
20. Purcarea VL. The impact of marketing strategies in healthcare systems. J Med Life 2019;12(2):93-96. [doi: 10.25122/jml-2019-1003] [Medline: 31406509]
21. Purdie-Vaughns V, Steele CM, Davies PG, Ditlmann R, Crosby JR. Social identity contingencies: how diversity cues signal threat or safety for African Americans in mainstream institutions. J Pers Soc Psychol 2008 Apr;94(4):615-630. [doi: 10.1037/0022-3514.94.4.615] [Medline: 18361675]
22. Herman JL, Flores AR, O'Neill KK. How many adults and youth identify as transgender in the United States? UCLA Williams Institute. 2022. URL: https://williamsinstitute.law.ucla.edu/publications/trans-adults-united-states [accessed 2024-12-13]
23. Hasenbush A, Flores AR, Kastanis A, Sears B, Gates GJ. The LGBT divide: a data portrait of LGBT people in the midwestern, mountain & southern states. UCLA Williams Institute. 2014. URL: https://williamsinstitute.law.ucla.edu/lgbtdivide/ [accessed 2024-12-13]
24. 2022 State Equality Index. Human Rights Campaign. 2023. URL: https://reports.hrc.org/2022-state-equality-index [accessed 2024-12-13]
25. Jennings L, Goût B, Whittaker PJ. Gender inclusive language on public-facing maternity services websites in England. Br J Midwifery 2022 Apr 2;30(4):208-214. [doi: 10.12968/bjom.2022.30.4.208]
26. Wu HY, Yin O, Monseur B, et al. Lesbian, gay, bisexual, transgender content on reproductive endocrinology and infertility clinic websites. Fertil Steril 2017 Jul;108(1):183-191. [doi: 10.1016/j.fertnstert.2017.05.011] [Medline: 28579417]
27. Weissman AL. Repronormativity and the reproduction of the nation-state: the state and sexuality collide. J GLBT Fam Stud 2016 Sep 9;13(3):277-305. [doi: 10.1080/1550428X.2016.1210065]

28. Green H, Riddington A. Gender inclusive language in perinatal services: mission statement and rationale. Brighton and Sussex University Hospitals. 2020. URL: https://www.bsuh.nhs.uk/maternity/wp-content/uploads/sites/7/2021/01/Gender-inclusive-language-in-perinatal-services.pdf [accessed 2023-07-16]

Abbreviations

HEI: Healthcare Equality Index
L&D: labor and delivery
LGBTQIA+: lesbian, gay, bisexual, transgender, queer, intersex, and asexual plus other gender- and sexual-oriented identities
TGNB: transgender and nonbinary

Edited by K Adapa; submitted 24.09.23; peer-reviewed by F Storey, GP Quinn; revised version received 13.07.24; accepted 30.07.24; published 06.01.25.

Please cite as:
Isaac SM, Dawes M, Howell ER, Oladipo AF
Gender-Inclusive Language in Public-Facing Labor and Delivery Web Pages in the New York Tristate Area: Cross-Sectional Study
JMIR Hum Factors 2025;12:e53057
URL: https://humanfactors.jmir.org/2025/1/e53057
doi:10.2196/53057

Leveraging Smart Telemedicine Technology to Enhance Nursing Care Satisfaction and Revolutionize COVID-19 Care: Prospective Cohort Study

You-Lung Chang[1*], MD; Chi-Ying Lin[1*], MD; Jiun Hsu[2], MD, PhD; Sui-Ling Liao[3], RN; Chun-Ti Yu[3], RN, HN; Hung-Chueh Peng[1], MBA; Chung-Yu Chen[1,4], MD, PhD; Matthew Huei-Ming Ma[1,4], MD, PhD; Juey-Jen Hwang[5], MD, PhD

[1]Department of Internal Medicine, National Taiwan University Hospital Yunlin Branch, Douliou, Yunlin County, Taiwan

[2]Medical Information Technology Office, National Taiwan University Hospital Yunlin Branch, Yunlin County, Taiwan

[3]Department of Nursing, National Taiwan University Hospital Yunlin Branch, Yunlin County, Taiwan

[4]College of Medicine, National Taiwan University, Taipei, Taiwan

[5]Department of Cardiovascular Medicine, Fu Jen Catholic University Hospital, New Taipei, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chung-Yu Chen, MD, PhD
Department of Internal Medicine, National Taiwan University Hospital Yunlin Branch, Douliou, Yunlin County, Taiwan

Abstract

Background: Telemedicine has been utilized in the care of patients with COVID-19, allowing real-time remote monitoring of vital signs. This technology reduces the risk of transmission while providing high-quality care to both self-quarantined patients with mild symptoms and critically ill patients in hospitals.

Objective: This study aims to investigate the application of telemedicine technology in the care of patients with COVID-19, specifically focusing on usability, effectiveness, and patient outcomes in both home isolation and hospital ward settings.

Methods: The study was conducted between January 2022 and December 2022. More than 800 cases were monitored using the QOCA remote home care system, a telemedicine platform that enables remote monitoring of physiological data—including heart rate, blood pressure, temperature, and oxygen levels—through Internet of Things devices and a 4G-connected tablet. Of these, 27 patients participated in thie study: the QOCA remote home care system was deployed 36 times in the isolation ward and 21 times to those in home isolation. The QOCA remote care system monitored isolated cases through remote care packages and a 4G tablet. Case managers and physicians provided telemedicine appointments and medications. Innovative methods were developed to enhance usage, including online health education, remote care equipment instructions via QR core links, and video consultations for patients without smartphones.

Results: A clinical nurse satisfaction survey revealed that most respondents found the content of the remote care package comprehensive and the interface easy to learn. They expressed a desire to continue using the system. The majority also agreed that using the remote care system and package would reduce their workload and that patients and caregivers could easily learn to use the package. While some respondents expressed concerns about network and Bluetooth connectivity, the majority (24/27, 89%) agreed to include the remote device as part of their routine equipment, with an average score of 84.8 points.

Conclusions: The integration of telemedicine technology improves the quality of care while reducing the workload and exposure of health care workers to viruses.

(JMIR Hum Factors 2025;12:e53456) doi:10.2196/53456

KEYWORDS

COVID-19; telemedicine; smart home care; intelligent medical care; nursing care; medical care; remote monitoring; vital signs; quarantine; home-care; hospital-care; screening; treatment; mHealth; mobile health; digital health; health education; remote equipment; smartphone; video consultation; remote care; caregiver; patient; quality of care; medical staff

Introduction

COVID-19, a global pandemic, has posed significant challenges to health care systems worldwide [1,2]. While Taiwan has achieved success in epidemic prevention, the reopening of borders raises concerns about potential gaps in prevention measures and strains on health care resources. Addressing these

challenges, this study identifies critical needs in the context of COVID-19 care.

The need for home quarantine and home medical care is vital. National Taiwan University Hospital Yunlin Branch has pioneered a telecommunication-based remote consultation model [3], providing video consultations for individuals under home quarantine. This initiative has not only met their basic health care needs but also minimized the risk of infection by reducing hospital visits. However, with the majority of cases being mild or asymptomatic, and a higher proportion of symptomatic cases among middle-aged and older individuals [4], it is crucial to develop a comprehensive telemedicine model for continuous medical care during home quarantine or isolation and for patients with mild COVID-19 undergoing home treatment [5-7].

During the COVID-19 pandemic, home monitoring of patients has gained popularity. There are two main research designs: (1) assessing patients before hospital admission to reduce virus exposure and identify deterioration and (2) postdischarge monitoring to provide continued care [8]. Monitoring methods include online platforms, paper-based operations with telephone interviews, or wearable devices [9-11]. Paper-based operations with telephone interviews are more inclusive due to limited digital literacy among the target population [12,13]. Most studies focus on vital signs rather than psychological issues. Wearing a pulse oximeter is important for monitoring, but it may cause anxiety and has limitations for certain patients [14-16]. Home monitoring during COVID-19 has benefits and can reduce health care costs [17]. Improving accessibility and convenience is key for future adoption.

Due to the pandemic, health care providers are unable to provide patients with the usual standard of care during hospitalization, leading to various challenges and impacts. Health care workers face psychological stress and staffing shortages, while patients experience isolation and mental health issues. Family members are unable to visit their loved ones in critical condition [18,19]. Technology can offer some assistance in addressing these issues, such as self-monitoring of blood oxygen levels and the integration of vital sign monitoring [20]. Collaboration between technology companies and health care providers is crucial, especially as the pandemic becomes normalized in the community. The development of information-assisted solutions is urgent in order to improve health care delivery [21].

There is a need for low-contact health care for hospitalized, isolated patients. To mitigate the risk of infection among patients with confirmed COVID-19 infection or those at high risk of being infected in dedicated isolation wards, the integration of smart ward solutions and telemedicine is proposed. Real-time remote physiological monitoring can minimize exposure risks and reduce the workload of health care staff, optimize management processes, and preserve valuable isolation rooms and personal protective equipment. Furthermore, obtaining real-time patient information enables early intervention and improves work efficiency, thereby enhancing the safety of patients with confirmed COVID-19 infection.

The study was conducted as a collaboration between National Taiwan University Hospital Yunlin Branch and Quanta Computer, utilizing their expertise in big data analysis and telecommunication telemedicine technology. The project focuses on the COVID-19 response hospitals, with the National Taiwan University Hospital Yunlin Branch serving as the testing ground. Specifically, it aims to provide assistance to patients with confirmed COVID-19 infection or those at high risk of being infected in the dedicated isolation ward, using advanced telemedicine equipment for clinical care. By implementing real-time remote physiological monitoring, the project aims to reduce the exposure risk and workload of health care personnel, simplify management processes, and optimize the utilization of valuable isolation rooms and personal protective equipment. Furthermore, by promptly assessing patient conditions and intervening as needed, the project seeks to enhance work efficiency and ensure the safety of patients with confirmed COVID-19 infection.

Methods

Ethical Considerations

The Institutional Review Board (IRB) of National Taiwan University Hospital approved this study (202009106RIPA). Questionnaires were collected after obtaining informed consent.

All participant data were anonymized or deidentified to ensure privacy. Participants voluntarily provided informed consent, and no personal identifiers were included in the analysis or dissemination. Additionally, anonymous surveys were used to ensure there were no risks of information leakage.

Participants in the study were not compensated monetarily or otherwise. Their involvement was voluntary, and all necessary support for participation, such as the provision of remote care packages and telemedicine tools, was offered free of charge. This ensured equitable participation without financial coercion.

Study Design, Data Sources, and Population

The study was conducted at the National Taiwan University Hospital Yunlin Branch. The inclusion criteria comprised individuals diagnosed with COVID-19, while exclusion criteria included moderate to severe COVID-19 cases, patients unwilling to participate in telemedicine monitoring, and individuals deemed unsuitable for telemedicine evaluation by physicians. Participants were required to be 20 years of age or older and provide informed consent approved by the IRB. The study enrollment period spanned from January to December 2022.

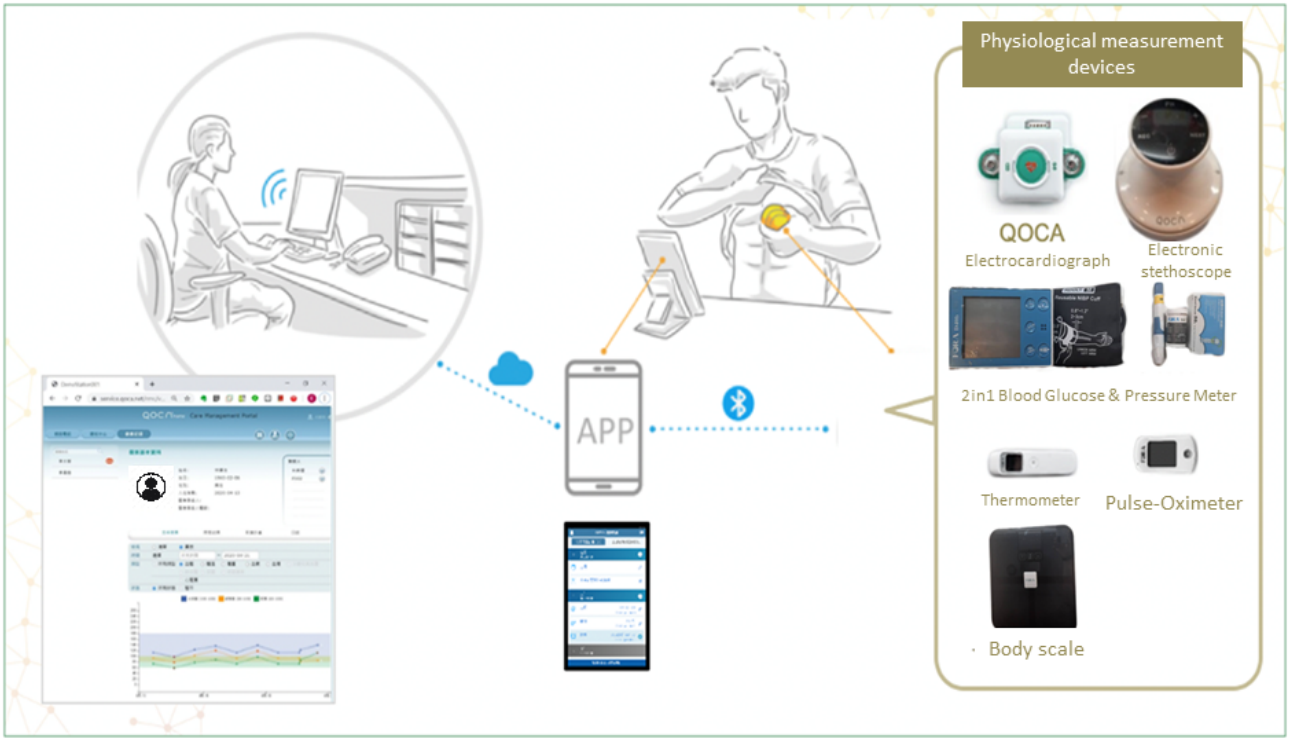
Home Quarantine and Home Medical Care

This study included individuals who have tested positive for COVID-19 and were classified as mild cases or have been determined by the Center for Disease Control, Taiwan, to not require immediate hospitalization due to severe conditions. These individuals can undergo self-isolation monitoring and were either required to quarantine or be monitored at home or quarantine hotels.

The QOCA home system, developed by Quanta Computer, is the core of the remote health care platform. It utilizes a 4G-connected tablet computer for patients, offering features like remote video communication, physiological signal measurement transmission, and remote physical examination (Figure 1). Home monitoring occurs at the patient's residence

or quarantine hotel, providing audiovisual communication and simple consciousness assessment. Vital signs, such as heart rate, blood pressure, temperature, and oxygen level, are measured and transmitted using Internet of Things (IoT) devices. A remote physical examination system, including electronic stethoscope devices, captures and uploads important findings like lung and heart sounds. An interactive care app records visual aspects and clinical conditions, which can be uploaded for health care providers. An online questionnaire system tracks COVID-19 symptoms, comorbidity control, and psychological well-being. Medication assessment and personalized care planning are provided. The hospital-side system manages home isolation quarantine and home-based COVID-19 care. Case managers track patient conditions through a dashboard, maintain daily health logs, and determine if hospital treatment is needed. Information and communication technology–based remote health care ensures continuous medical care and psychological support for individuals in home quarantine or isolation and home-based patients with COVID-19. Remote care packages enable self-isolated individuals to monitor and report symptoms and vital signs, with data transmitted back to the hospital. Dedicated care managers communicate with patients, report to physicians, and provide necessary support in case of discomfort or abnormalities.

Figure 1. The QOCA home system by Quanta Computer is central to the remote health care platform. Utilizing a 4G tablet, it enables remote communication, physiological data transmission, and virtual exams. Monitoring occurs at homes or quarantine locations, allowing audiovisual contact and basic assessment. IoT devices measure and send vital signs, while an electronic stethoscope captures and uploads lung and heart sounds. An interactive app records visual and clinical data for health care providers. An online questionnaire tracks COVID-19 symptoms, comorbidity control, and psychological well-being and includes medication assessment and personalized care planning. IoT: Internet of Things.



Smart Hospital Ward

Patients with COVID-19 admitted to negative pressure isolation wards or intensive care units at National Taiwan University Hospital Yunlin Branch were monitored using the QOCA home system developed by Quanta Computer. Tablet computers, IoT physiological monitoring devices, and video cameras were installed in patient rooms for remote monitoring. Patient physiological data, including heart rate, respiration, body temperature, blood pressure, and blood oxygen levels, were automatically recorded and uploaded to the cloud. The care team accessed these data without entering the rooms. The smart ward included an electronic stethoscope device for recording heart and lung sounds. A central monitoring module and a patient dynamic dashboard were established at the nursing station, providing alerts for abnormal conditions. Patient laboratory reports and images were integrated, enabling health care providers to interact with patients through the video system. The study aimed to utilize information and communication technology, including 4G-connected tablets and IoT monitoring platforms, to facilitate remote health care for patients with COVID-19 in isolation units. This system enabled continuous medical care, psychological support, and remote consultations with physicians; nursing care; psychological counseling; and rehabilitation therapy. It allowed for monitoring of vital signs and COVID-19 symptoms, promoting patient well-being without direct contact.

Satisfaction Survey for the QOCA Smart Ward Remote Care System: Clinical Nursing Staff and General Population

This questionnaire aimed to understand the relevant settings of the remote care system and its clinical application in dedicated wards for COVID-19. The information gathered will be used for future improvements and optimization. This survey was anonymous, and there were no concerns about personal

information leakage. The questionnaire included aspects of communication quality, user-friendliness, overall satisfaction, and suggestions. Each option consisted of strongly agree, agree, neutral, disagree, and strongly disagree (Multimedia Appendix 1).

Results

Overview

Before officially enrolling patients in the monitoring care package, thorough explanations and assistance were provided to each individual in the isolation ward to ensure their understanding and obtain their consent to participate in the research study. Following this, every isolated patient was equipped with a set of remote care packages. This became standard practice in the ward, significantly reducing the frequency of health care personnel entering patient rooms to don and remove isolation garments. Patients were empowered to self-measure their vital signs whenever they felt unwell, providing immediate access to physiological data that could alert health care professionals to any discomfort or changes in their condition. This approach not only alleviated patients' psychological anxiety about their disease progression but also allowed health care staff to spend less time in isolation garments while maintaining a real-time understanding of the patients' current status, ultimately improving the quality of clinical care.

Enhancing Hardware Infrastructure for Remote Care System in Isolation Ward

To enhance the hardware aspect of the remote care system implemented in the isolation ward, the original 7-bed negative pressure ward was expanded to accommodate the installation of 32 sets of care packages, with one set allocated to each bed. Additionally, the nursing treatment cart and the nursing station computer were equipped with the QOCA remote care system platform and individual nursing staff accounts. This setup empowered the nursing personnel to monitor the patients' vital sign data in real time, which were automatically uploaded by the system after patient measurements. Continuous monitoring of vital signs was made possible throughout the patients' entire hospitalization and isolation period. Moreover, the QOCA system facilitated real-time video communication when necessary, enabling consultations with psychiatric physicians to address any illness-related anxieties that the patients may have. By extending the focus of care beyond the patients' physiological needs, the hospitalization and treatment process also addressed their psychological well-being.

Remote Home Care Package for Individual Who Are COVID-19 Positive Under Self-Isolation

For individuals who have tested positive for COVID-19 and can self-isolate at home without requiring hospitalization, a comprehensive remote home care package was provided. This package enabled individuals to independently monitor their health, even without monitoring devices at home. It included daily self-observation of symptoms and measurement of vital signs, with the numerical data automatically transmitted to the hospital through a tablet computer connected to 4G.

To ensure prompt attention to any discomfort or adverse conditions, a dedicated care manager was assigned to each patient. The care manager promptly reported any concerns to the responsible physician and maintained regular contact to inquire about the patient's condition throughout the home isolation period. If individuals experience symptoms, they could also schedule telemedicine appointments to closely monitor changes in their condition.

In cases where symptom management was necessary, the responsible physician could prescribe appropriate medications to help alleviate symptoms. This comprehensive remote home care package aimed to provide effective care and support to individuals with COVID-19, ensuring their well-being while minimizing the risk of further transmission.

Advancing Care Through QOCA Telecare System for COVID-19 Monitoring and Isolation

The number of cases monitored using the QOCA remote home care system exceeded 800. Among them, 29 individuals voluntarily participated in this research study and provided their consent by signing the IRB consent form. The QOCA remote home care system was deployed 36 times in the isolation ward and 21 times to those in home isolation, as indicated in Table 1. In the ward, patients primarily monitor themselves or receive assistance from family members for monitoring, while a small number of individuals without family support or hired caregivers receive help from nursing staff for measurements.

The participants in this study represented a diverse range of age groups, with the majority being aged 41-60 years (13/29, 45%), followed by the 61-80 years age group (9/29, 31%), and the fewest participants in the 21-40 years (3/29, 10%) and 80+ years age groups (4/29, 14%). The sample was predominantly male, with 27 (93%) male participants and only 2 (7%) female participants. Regarding the need for assistance, 11 (38%) participants required help during the study, while 18 (62%) managed independently. The participants' education levels varied, with most having elementary or junior high education (20/29, 69%), while a smaller number had completed high school or higher education (9/29, 31%). In terms of technology usage, 21 (72%) participants used digital devices daily, while 8 (28%) reported no prior experience. Health engagement levels were mixed, with the majority (20/29, 69%) not actively seeking information about their conditions, while some participants regularly consumed health-related content. Additionally, 17 (59%) participants had prior experience with health monitoring devices, further illustrating the range of familiarity with telecare technology among the group.

In contrast to traditional inpatient monitoring methods, the ward implemented a specialized remote home care system designed specifically for the isolation care required during the COVID-19 pandemic. This system enables patients to independently measure their vital signs and automatically transmit the data to the nursing staff. To ensure effective utilization of the remote home care system by patients and their accompanying family members, innovative care methods have been developed, including the following:

1. Online health education materials: With the widespread use of smartphones due to technological advancements, disease-specific health education materials are provided through QR codes during hospitalization. This allows patients to access relevant information whenever needed.
2. Guidelines for operating remote care devices: Some individuals may be unfamiliar with monitoring devices. Even after receiving instructions from health care professionals, patients may require repeated practice. To facilitate practice sessions without overwhelming the nursing staff, instructional videos demonstrating device operation can be accessed through QR codes. Patients can practice as many times as necessary to ensure accurate measurements.
3. Tablet-based video consultations for remote care: Patients without smartphones can still access important information through tablet-based video consultations offered by the remote care system.

The ward continues to fully utilize the remote home care system for clinical monitoring, and the number of users continues to grow. Health care professionals not only provide personalized care to patients but also make adjustments to the care process based on the implementation of the remote home care system. This approach reduces the risk of viral exposure and eases the workload for health care staff, streamlines management procedures, and ultimately enhances the quality of care provided.

Table . Results of case enrollment.

	Value
QOCA remote home care system, n	
Isolation ward	36
Home isolation	21
Total	57
Goal	50
Total utilization (cases), n	>800
Execution rate	100%

Clinical Nursing Staff Satisfaction Survey

This survey aimed to assess the satisfaction and experiences of the clinical nursing staff in using the telecare system in the dedicated COVID-19 isolation wards. The questionnaire was divided into 3 main categories: *communication quality*, *user-friendliness*, and *overall experience*. The objective was to identify and analyze the challenges and difficulties faced by users in relation to equipment, signals, and interface operations.

A total of 27 questionnaires were collected exclusively from nursing staff working in the dedicated COVID-19 wards, specifically Ward 7A and Ward 7B. Regarding the transmission signals of the equipment, 44% (n=12) of the respondents believed that the internet connection of the telecare system was stable, while 37% (n=10) found the Bluetooth connection between the tablet and the devices in the package to be stable. When using the telecare system for video calls, 56% (n=15) stated that the image quality was clear, and 56% (n=15) stated that the sound quality was clear (Table 2).

In terms of usability, following training on the telecare system's interface, 89% (n=24) of the respondents found it easy to learn, and an equal percentage (n=24, 89%) expressed their willingness to continue using the telecare system to assist in clinical care. Furthermore, 67% (n=18) believed that patients or caregivers could quickly grasp the usage of the telecare package. Moreover, 81% (n=22) of the respondents indicated that the telecare package contained all the necessary components and fulfilled the requirements for clinical monitoring. Additionally, 89% (n=24) agreed that the use of the telecare system and package could help alleviate the burden of clinical care (Table 3).

Overall, 89% (n=24) of the respondents agreed to incorporate the telecare device as a regular piece of equipment in isolation wards, with an average satisfaction score of 84.8. The majority (24/27, 89%) of the responses highlighted the short monitoring time of the device and the instability of signal transmission. They expressed a desire for improved stability in wireless internet and Bluetooth signals, as it would lead to a smoother user experience (Table 4).

Table . Communication quality survey responses regarding the telemedicine system.

Question	Score[a] (n=27), n (%)				
	1	2	3	4	5
I think the network connection of this telecare system is stable	1 (4)	11 (41)	12 (44)	3 (11)	0 (0)
The Bluetooth connection between the tablet in the care bag and the instrument is stable	2 (7)	8 (30)	12 (44)	5 (19)	0 (0)
When using the nursing system, the image is very clear.	3 (11)	12 (44)	6 (22)	5 (19)	1 (4)
When using the nursing system, the sound is very clear.	2 (7)	13 (48)	8 (30)	3 (11)	1 (4)

[a]Score: 1=strongly agree, 2=agree, 3=neutral, 4=disagree, and 5=strongly disagree.

Table . User-friendliness survey responses regarding the telemedicine system.

Question	Score[a] (n=27), n (%)				
	1	2	3	4	5
The interface of this telecare system is easy for me to use after being taught.	15 (56)	9 (33)	1 (4)	1 (4)	1 (4)
I find the equipment in this telecare package difficult to use.	1 (4)	3 (11)	1 (4)	6 (22)	16 (59)
I think most patients or caregivers can quickly learn to use the telecare package.	7 (26)	11 (41)	4 (15)	3 (11)	2 (7)
I think the contents of this telecare package are complete and meet the needs of clinical monitoring	12 (44)	10 (37)	3 (11)	1 (4)	1 (4)

[a]Score: 1=strongly agree, 2=agree, 3=neutral, 4=disagree, and 5=strongly disagree.

Table . Overall satisfaction with the QOCA remote care system in smart wards.

Question and score	Value (n=27), n (%)
I think the use of telecare systems and care packages can help reduce the load of clinical care.[a]	
1	15 (56)
2	9 (33)
3	2 (7)
4	0 (0)
5	1 (4)
I would like to continue to use this telecare system to assist clinical care.[a]	
1	15 (56)
2	9 (33)
3	2 (7)
4	0 (0)
5	1 (4)
I think the remote care system and care package can become one of the routine equipment used by the unit.[a]	
1	15 (56)
2	6 (22)
3	4 (14)
4	1 (4)
5	1 (4)
If a full score is 100, how much would you like to rate the telecare system and care package?	
60	1 (4)
70	2 (7)
75	1 (4)
80	6 (22)
85	3 (11)
90	11 (41)
95	2 (7)
99	1 (4)

[a]Score: 1=strongly agree, 2=agree, 3=neutral, 4=disagree, and 5=strongly disagree.

Discussion

Principal Findings

The enhancement of the hardware infrastructure for the remote care system in the isolation ward has significantly improved the quality of care and patient outcomes [22,23]. By expanding to accommodate more care packages, each patient receives continuous monitoring and support. The integration of the QOCA remote care system into nursing carts and computers allows real-time access to vital sign data, enabling prompt interventions and ensuring patient well-being. The system also facilitates video consultations with psychiatric physicians, addressing patients' anxieties and holistic needs. This approach benefits both patients and nursing staff, who can efficiently monitor multiple patients, respond promptly to changes, and access patient data remotely, thus saving time and reducing errors.

The implementation of a remote home care package for individuals who are COVID-19 positive in self-isolation offers effective care and support while reducing transmission risk [24]. Patients can independently monitor their health by observing symptoms and measuring vital signs, actively participating in their care. A tablet connected to 4G enables real-time transmission of data to health care providers, facilitating informed decisions. Dedicated care managers provide personalized support, regularly checking in on patients and promptly reporting concerns. Telemedicine appointments allow close monitoring, adjustment of treatment plans, and direct access to medical professionals, thus reducing anxiety. Remote prescription of medications ensures symptom management without visiting a health care facility. However, challenges include subjective self-reporting and disparities in internet access and technological skills [23,25]. Education and efforts to address disparities are essential for equitable care.

The implementation of the QOCA remote home care system for COVID-19 monitoring and isolation has been successful in advancing care and improving patient outcomes. With over 800 cases monitored, the system allows patients to independently measure vital signs and transmit data to nursing staff, reducing the need for frequent in-person monitoring. Innovative methods, such as online health education materials and instructional videos, ensure optimal utilization of the system. Tablet-based video consultations cater to patients without smartphones. By fully utilizing the remote home care system, the ward provides personalized care, streamlines management procedures, reduces viral exposure, and eases the workload for staff, ultimately enhancing care quality [26]. Ongoing monitoring and evaluation are crucial for system refinement, and its success in COVID-19 care opens doors for its application in other health care settings, showcasing the transformative potential of technological advancements in patient care.

During the COVID-19 pandemic, nursing professionals have made substantial achievements in delivering adequate patient care by utilizing telehealth [8,12,24]. The results of the clinical nursing staff satisfaction survey regarding the telecare system in COVID-19 isolation wards provide valuable insights into their experiences and challenges [27]. The majority of respondents found the telecare package to be usable and to have met the requirements for clinical monitoring. After training, the respondents stated that the interface was easy to learn, and they expressed their willingness to continue using the system. A substantial percentage believed that patients and caregivers could quickly grasp its usage, and that the use of the telecare system helped alleviate the burden of clinical care [12]. Regarding transmission signals, opinions were mixed. While a portion of respondents found the internet connection stable, others highlighted issues with Bluetooth connectivity. In terms of video calls, a majority found the image and sound quality to be clear. Overall, the survey indicates high satisfaction with the telecare system, with a majority of respondents expressing their support for incorporating it as a regular equipment in isolation wards. The average satisfaction score was positive [13]. However, feedback emphasized the need for improvements in signal stability, particularly in wireless internet and Bluetooth connections [24]. The short monitoring time of the device was also a concern.

The difficulties and obstacles in telemedicine for COVID-19 care include insufficient hardware environment and the short monitoring time of equipment [28]. The negative pressure isolation environment in the ward, with multiple walls and controlled doors, leads to weak signals. Installing signal amplifiers, as suggested by the hospital's IT department, can improve signal stability. The monitoring time of the equipment is too short, but through discussions with engineers and the manufacturer, it is believed that adjusting the monitoring time according to individual patient needs can align it with clinical requirements, enabling more accurate monitoring of values. Addressing these challenges will enhance the effectiveness and reliability of telemedicine in COVID-19 care.

One potential limitation of this study is the relatively small sample size, which may affect the generalizability of the findings. With only 36 deployments in the isolation ward and 21 deployments in home isolation, the results may not fully represent the broader population of patients with COVID-19 or other health care contexts. Additionally, the voluntary nature of participation could introduce selection bias, as those who agreed to participate may differ from those who declined in terms of health status or access to technology. Another limitation is that the study was conducted within a single hospital, potentially limiting the applicability of the results to other settings with different resources or patient demographics. Lastly, the short-term focus of the study may not capture the long-term effectiveness or sustainability of telemedicine interventions, warranting further research in this area.

Conclusion

The implementation of the QOCA telemedicine system demonstrated significant benefits in managing patients with COVID-19 who have mild symptoms, both in isolation wards and home quarantine settings. By enabling continuous remote monitoring of vital signs and offering real-time communication between patients and health care providers, the system not only improved the quality of care but also reduced the workload and exposure risk for medical staff.

However, the study had some limitations, including a relatively small sample size and the potential for selection bias, which may affect the generalizability of the findings. Additionally, the short-term nature of the study may not fully capture the long-term impacts of telemedicine in patient care.

Despite these limitations, our findings suggest that telemedicine systems like QOCA can play a crucial role in pandemic response and future health care delivery. Further research with larger, more diverse populations and extended follow-up periods is recommended to fully explore the potential of telemedicine in improving patient outcomes and health care efficiency.

Acknowledgments

We sincerely thank Quanta Computer for their invaluable support in establishing the following health care systems: the Smart Ward Cloud Care System, QOCA Home, Intelligent Remote Health Care System, and Remote Home Care Package. This study was supported by the National Science and Technology Council, Taiwan, for the research (MOST 109－2823－8－002－013-CV).

Data Availability

All data generated or analyzed during this study are included in this published paper and its supplementary materials.

Authors' Contributions

Conceptualization: MHMM, JJH
Data curation: YLC, CYL
Investigation: JH, SLL, CTY, HCP
Methodology: JH, SLL, CTY, HCP, MHMM, JJH
Software: JH
Supervision: CYC, MHMM, JJH
Validation: CYC
Visualization: JH, SLL, CTY
Writing – original draft: YLC, CYL
Writing – review & editing: CYC

Conflicts of Interest

None declared.

References

1. Mahase E. COVID-19: WHO declares pandemic because of “alarming levels” of spread, severity, and inaction. BMJ 2020 Mar 12;368:m1036. [doi: 10.1136/bmj.m1036] [Medline: 32165426]
2. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/NEJMoa2001017] [Medline: 31978945]
3. Chang KC, Hung SH, Hsueh SJ, et al. Development of the Houston-Apollo model for older people living in remote areas in Taiwan. Geriatr Gerontol Int 2021 Mar;21(3):299-305. [doi: 10.1111/ggi.14130] [Medline: 33527540]
4. Zhou F, Yu T, Du R, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062. [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
5. Hollander JE, Carr BG. Virtually Perfect? telemedicine for COVID-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/NEJMp2003539] [Medline: 32160451]
6. COVID-19 telehealth program (invoices & reimbursements). Federal Communications Commission. URL: https://www.fcc.gov/covid-19-telehealth-program [accessed 2021-12-12]
7. Telehealth.HHS.gov. US Department of Health and Human Services. URL: https://www.hhs.gov/coronavirus/telehealth/index.html [accessed 2021-12-12]
8. Vindrola-Padros C, Singh KE, Sidhu MS, et al. Remote home monitoring (virtual wards) for confirmed or suspected COVID-19 patients: a rapid systematic review. E Clin Med 2021 Jul;37:100965. [doi: 10.1016/j.eclinm.2021.100965] [Medline: 34179736]
9. O’Keefe JB, Tong EJ, Taylor TH Jr, O’Keefe GAD, Tong DC. Use of a telemedicine risk assessment tool to predict the risk of hospitalization of 496 outpatients with COVID-19: retrospective analysis. JMIR Public Health Surveill 2021 Apr 30;7(4):e25075. [doi: 10.2196/25075] [Medline: 33667174]
10. Margolius D, Hennekes M, Yao J, et al. On the front (phone) lines: results of a COVID-19 hotline. J Am Board Fam Med 2021 Feb;34(Suppl):S95-S102. [doi: 10.3122/jabfm.2021.S1.200237] [Medline: 33622824]
11. Hutchings OR, Dearing C, Jagers D, et al. Virtual health care for community management of patients with COVID-19 in Australia: observational cohort study. J Med Internet Res 2021 Mar 9;23(3):e21064. [doi: 10.2196/21064] [Medline: 33687341]
12. Greenhalgh T, Knight M, Inda-Kim M, Fulop NJ, Leach J, Vindrola-Padros C. Remote management of COVID-19 using home pulse oximetry and virtual ward support. BMJ 2021 Mar 25;372:677. [doi: 10.1136/bmj.n677] [Medline: 33766809]
13. Agarwal P, Mukerji G, Laur C, et al. Adoption, feasibility and safety of a family medicine-led remote monitoring program for patients with COVID-19: a descriptive study. CMAJ Open 2021 Apr 1;9(2):E324-E330. [doi: 10.9778/cmajo.20200174] [Medline: 33795222]
14. COVID-19 clinical management: living guideline. World Health Organization. 2021 Jan 25. URL: https://apps.who.int/iris/bitstream/handle/10665/338882/WHO-2019-nCoV-clinical-2021.1-eng.pdf [accessed 2025-01-06]
15. Nunan J, Clarke D, Malakouti A, et al. Triage Into the Community for COVID-19 (TICC-19) patients pathway - service evaluation of the virtual monitoring of patients with COVID pneumonia. Acute Med 2020;19(4):183-191. [Medline: 33215171]
16. Feiner JR, Severinghaus JW, Bickler PE. Dark skin decreases the accuracy of pulse oximeters at low oxygen saturation: the effects of oximeter probe type and gender. Anesth Analg 2007 Dec;105(6 Suppl):S18-S23. [doi: 10.1213/01.ane.0000285988.35174.d9] [Medline: 18048893]
17. Poncia HD, Ryan J, Carver M. Next day telephone follow up of the elderly: a needs assessment and critical incident monitoring tool for the accident and emergency department. J Accid Emerg Med 2000 Sep;17(5):337-340. [doi: 10.1136/emj.17.5.337] [Medline: 11005403]

18. Deressa W, Worku A, Abebe W, Gizaw M, Amogne W. Risk perceptions and preventive practices of COVID-19 among healthcare professionals in public hospitals in Addis Ababa, Ethiopia. PLoS ONE 2021 Jun 25;16(6):e0242471. [doi: 10.1371/journal.pone.0242471] [Medline: 34170910]
19. Wakam GK, Montgomery JR, Biesterveld BE, Brown CS. Not dying alone - modern compassionate care in the COVID-19 pandemic. N Engl J Med 2020 Jun 11;382(24):e88. [doi: 10.1056/NEJMp2007781] [Medline: 32289215]
20. Sutherland AE, Stickland J, Wee B. Can video consultations replace face-to-face interviews? palliative medicine and the COVID-19 pandemic: rapid review. BMJ Support Palliat Care 2020 Sep;10(3):271-275. [doi: 10.1136/bmjspcare-2020-002326] [Medline: 32457086]
21. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 2;6(2):e18810. [doi: 10.2196/18810] [Medline: 32238336]
22. Shen YT, Chen L, Yue WW, Xu HX. Digital technology-based telemedicine for the COVID-19 pandemic. Front Med (Lausanne) 2021 Jul 6;8:646506. [doi: 10.3389/fmed.2021.646506] [Medline: 34295908]
23. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
24. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 1;20(1):1193. [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
25. Coffey JD, Christopherson LA, Williams RD, et al. Development and implementation of a nurse-based remote patient monitoring program for ambulatory disease management. Front Digit Health 2022 Dec 14;4:1052408. [doi: 10.3389/fdgth.2022.1052408] [Medline: 36588748]
26. Bardhan R, Byrd T, Boyd J. Workforce management during the time of COVID-19—lessons learned and future measures. COVID 2023;3(1):1-27. [doi: 10.3390/covid3010001]
27. Wu Y. Utilization of telehealth and the advancement of nursing informatics during COVID-19 pandemic. Int J Nurs Sci 2021 Oct 10;8(4):367-369. [doi: 10.1016/j.ijnss.2021.09.004] [Medline: 34552807]
28. Zhang T, Mosier J, Subbian V. Identifying barriers to and opportunities for telehealth implementation amidst the COVID-19 pandemic by using a human factors approach: a leap into the future of health care delivery? JMIR Hum Factors 2021 Apr 9;8(2):e24860. [doi: 10.2196/24860] [Medline: 33779566]

Abbreviations

IoT: Internet of Things
IRB: institutional review board

Edited by E Borycki; submitted 07.10.23; peer-reviewed by CY Chang, S Setia; revised version received 02.10.24; accepted 09.12.24; published 21.01.25.

Please cite as:
Chang YL, Lin CY, Hsu J, Liao SL, Yu CT, Peng HC, Chen CY, Ma MHM, Hwang JJ
Leveraging Smart Telemedicine Technology to Enhance Nursing Care Satisfaction and Revolutionize COVID-19 Care: Prospective Cohort Study
JMIR Hum Factors 2025;12:e53456
URL: https://humanfactors.jmir.org/2025/1/e53456
doi:10.2196/53456

Original Paper

Examining Individuals' Use of the Internet for Health Care Activities Over Time: Results from the US National Health Interview Survey

Zachary Junkins[1], MS; Nusrath Zahan[1], BS; David Neyens[1], MPH, PhD

Department of Industrial Engineering, Clemson University, Clemson, SC, United States

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: Telehealth is an increasingly important component of health care services. Telehealth services may present an opportunity to increase the equity, accessibility, and effectiveness of health care. As such, it is critical that telehealth design focuses on reducing the barriers to access and usability that may impair some telehealth users.

Objective: Our goal was to identify different demographic characteristics, behaviors, or opinions that may predict groups who are likely to face a barrier to using telehealth services.

Methods: We used data from the National Health Interview Survey and multiple logit regression models focused on different aspects of telehealth to examine three different avenues of telehealth service: looking up health information using the internet, scheduling an appointment using the internet, and communicating with a care provider through email using the internet in order to consider the ways in which different telehealth services may face different barriers.

Results: Our results suggest that middle-aged (36-55 years old) and older adult (56-85 years old) respondents were significantly less likely to look up health information using the internet or schedule an appointment using the internet versus younger individuals (18-35 years old). Specifically, our analysis found that middle-aged adults were found to have a higher odds ratio than older adults (0.83 vs 0.65) for looking up health information using the internet. We also found that there were differences in age groups for using technology to perform health care–related tasks. In terms of searching for health information using the internet and scheduling appointments using the internet, we found differences between men and women, with women being significantly more likely than men to look up health information using the internet, schedule an appointment using the internet, and communicate with a care provider through email using the internet. Across all the investigated variables, we found that the rates of using the internet for looking up health information, scheduling an appointment, and communicating with a care provider over email increased substantially across the study period. The impact of costs was inconsistent across the different models in our analysis. We also found that there is a strong correlation between respondents' collaboration in their personal health and the likelihood that they would use telehealth services to meet these needs.

Conclusions: This analysis provides an exploratory look at the data to highlight barriers that may impact a user's ability to access telehealth services in the context of other potential predictor variables to account for the real-world variability that these may present. Future work should examine the complex relationships of those variables and understand how these interactions are correlated with the respondents' use of telehealth.

(JMIR Hum Factors 2025;12:e58362) doi:10.2196/58362

KEYWORDS

internet; web search; internet search; internet use; searching behavior; access to health information; telemedicine; telehealth; virtual care; virtual health; virtual medicine; logistic regression model; regression model; National Health Interview Survey; NHIS

Introduction

Background

With the advancement of digital technologies, the use of technology in health care has grown [1,2]. While there have been several attempts to better understand the use of digital technologies in health care, much of the work has been done in silos and is sometimes limited in scope [2,3]. Generally, telehealth refers to health care communication through technology, often used in conjunction with telemedicine and eHealth [3-5]. Mobile health is an additional term that describes telehealth services in the context of mobile devices [6]. Telehealth is increasingly being used in growing populations, offering distinct advantages and potential barriers for different patient populations and their care.

Benefits Associated With Telehealth

The increased adoption of telehealth services has allowed for its impacts to be examined on actual patients receiving telehealth care [7]. It has been shown that when telehealth is implemented with current best practices, it may help to balance the health care supply and demand disparity [8], improve patient access to care [8,9], and reduce the cost of care [8,9]. Due to increasing demand with changes in population sizes and demographics, as well as decreasing supply as care providers retire or change careers [8], telehealth services and telehealth programs have helped bridge this gap [10,11]. Telehealth also has been shown to improve patients' ability to access care, thus enabling patients to receive efficient and cost-effective care [8,10] by reducing the impacts of geographical barriers [12] and cost barriers to care [13]. Telehealth may also improve the cost-effectiveness of care, with [14] finding that the "all-cause" cost of providing care to older patients (older than 65 years) decreased from US $937.25 to US $491.52 with the inclusion of telehealth services. Telehealth can also offer support for individuals seeking mental health care where mental health specialties are distant or when there may be patient privacy issues in obtaining mental health care [15].

Barriers to the Use of Telehealth Services

There are several potential barriers that have been identified related to the use and effective implementation of telehealth services [7]. A notable barrier to telehealth services is patients' health literacy [16-18]. Individuals with lower levels of health literacy have shown lower comfort levels with using telehealth technologies [17,19]. Older people are using as well as providing telehealth services [20]. Yet, it has been suggested that older patients (older than 60 years) might have greater difficulty using telehealth services [21]. Additionally, technological infrastructure issues (low-quality, limited internet access) may reduce the ability of individuals to access telehealth services [22]. Individuals' access to electronic devices (ie, smartphones, computers, and tablets) may also be a barrier to engaging in telehealth services [22]. In addition to the access issues, the time and costs required to engage in the services (both for the patients as well as clinicians) may be a barrier to rolling out large programs or integrating them into the workflow processes for clinical staff [23].

Changes in Health Care Populations Over Time

There are several factors that have been shown to impact the use of web-based resources for health care and other telehealth services [24]. A 2018 study suggests that demographic factors, such as living in rural areas, age, and insurance types, impact telehealth implementation and use in health care facilities [24]. There have also been racial and ethnic discrepancies reported in the literature, with Black and Hispanic patients preferring emergency departments to telemedicine compared to White patients [25]. Age is also an important factor as the use of telehealth differs among different age groups [14,25]. In addition, health insurance access, and having internet access also influence the use of telehealth [14].

It is important to consider how user needs with telehealth change over time [26]. Changing population demographics, such as age, have major impacts on the use and function of health care [27]. Other additional factors, such as socioeconomic factors and changes in the prevalence of different health conditions, also impact the function of the health care system [27]. Over time, population demographics, access, and technology can change significantly, leading to incorrect conclusions or missed critical trends due to observing only one moment. It is important to consider all these elements when examining which predictors may indicate barriers to telehealth access. The objective of this research is to examine how individuals used the internet to support their health care over 7 years (2012-2018) using the National Health Interview Survey (NHIS) data. Specifically, we examined how different demographic variables and respondents' perspectives influenced how individuals looked up health information using the internet, scheduled appointments using the internet, and emailed to communicate with a care provider using the internet.

Methods

Ethical Considerations

An ethics board review was not required for this analysis as this study used publicly available data.

Overview

Each year the Centers for Disease Control and Prevention (CDC), a government agency assigned to monitor the general public's behavior and health status conducts the NHIS, which is an annual cross-sectional survey designed to gather data on a variety of health-related topics throughout the United States with oversampling of certain demographic groups in a way that is nationally representative [28]. The survey weighting variable is poststratified based on the US Census data to represent national population characteristics [28]. These survey weights are identified for all combinations of persons within the variable WTFA_SA in the NHIS data. Due to the complex survey sampling strategy, the survey person weights must be used in analyzing the data to avoid substantial bias in the results [28].

For this analysis, the NHIS sample adult files for each year from 2012 to 2018 were used. For the files from 2012 to 2014, the American Standard Code for Information Interchange data file was combined using the stringr, stringi, foreign, and RCurl packages in R (R Foundation for Statistical Computing) to run

the associated Statistical Analysis System statements to produce a .csv version of the sample adult file. For the years 2015-2018, a csv file was provided for the sample adult survey data. We created a single file by combining responses for each year using the R functions rbindlist from the data.table package to match each survey year on the corresponding variable columns. Only identical questions related to our research objectives were included in the combined data set.

All responses recorded as the option "not ascertained" were recoded as NA for all variables. All the variables were recoded by assigning binary values or group responses to facilitate the data analysis. The variables used as outcome variables were recoded to become binary variables: health information seeking using the internet (HIT1A), scheduling health care appointments using the internet (HIT3A), and communication with health care providers through email using the internet (HIT4A). Demographic variables were also recoded, including the respondent's race (RACERPI2), Hispanic ethnicity (HISPANI_I), sex (SEX), region (REGION), marital status (R_MARITL), internet use (AWEBUSE) and the frequency of using the internet (AWEBOFNO and AWEBOFTP), email use (AWEBEML) and the frequency of using email (AWEBMTP), whether a respondent would go to a clinic or doctor's office (AUSUALPL and APLKIND), and where a respondent goes to seek preventative care (AHCPLKND). Additionally, whether a respondent skipped medication doses to save money (ARX12_1), whether a respondent took less medication to save money (ARX12_2), if a respondent had delayed filling a prescription (ARX12_3), preferred low-cost medication (ARX12_4), reported buying prescription drugs from another country to save money (ARX12_5), used alternative therapies to save money (ARX12_6), affordability of prescribed medication (AHCAFYR1), affordability of mental health care or counseling (AHCAFYR2), affordability of dental care (AHCAFYR3), affordability of eyeglasses (AHCAFYR4), and worries about paying medical bills (AWORPAY).

All responses where the data were coded as "not ascertained" or "missing" were recoded as "NAs" for all of the variables and were dropped from the analysis. Whether or not a survey respondent had looked up health information using the internet in the last 12 months was recoded to a binary response of 1 to indicate if the respondent had looked up health information on the internet or 0 if they did not indicate looking up health information on the internet (all other responses). Similarly, all binary variables were recoded in the same manner. A survey respondent's race was identified within the data by the RACERPI2 variable from the NHIS data. The variable was recoded to indicate if the survey respondent reported their race as being White, Black or African American, American Indian or Alaska Native, Asian, or multiple races. A survey respondent's Hispanic ethnicity was recoded to indicate if the survey respondent reported being Hispanic, multiple Hispanic, Puerto Rican, Mexican, Mexican-American, Cuban or Cuban American, Dominican (Republic), Central or South American, other Latin American (type not specified), other Spanish, Hispanic or Latino or Spanish (nonspecific type), Hispanic or Latino or Spanish (type refused), or were not Hispanic (a response of 12, not Hispanic or Spanish origin). A survey respondent's region was identified within the data using the variable REGION from the NHIS data. The variable was recoded to indicate if the survey respondent reported residing in the Northeastern, Midwestern, Southern, or Western United States. A survey respondent's marital status was recoded to indicate if the survey respondent reported living with a spouse or partner or not living with a spouse or partner (all other responses). A survey respondent's internet use was identified in the data using the AWEBUSE variable from the NHIS data. The variable was recoded to indicate if the survey respondent reported using the internet, or not. Additionally, a survey respondent's internet use frequency was identified in the data using the AWEBOFNO and AWEBOFTP variables. The variables were recoded to indicate frequently using the internet (responses of once per day or more frequently) or not using the internet frequently (all other response combinations). The variable APLKIND was recoded to indicate if a respondent goes to a clinic or doctor's office (a response of clinic or health center, or doctor's office or health maintenance organization) or somewhere other than a clinic or doctor's office (all other responses). These variables were combined to indicate if a respondent goes to a clinic or doctor's office when sick, goes somewhere other than a clinic or a doctor's office when sick, or does not indicate going anywhere when sick. Where a respondent goes to seek preventative care was identified using the variable AHCPLKND in the NHIS data. The variable was recoded to indicate if a respondent goes to a clinic or doctor's office for preventative care (a response of clinic or health center, or doctor's office or health maintenance organization) or does not go to a clinic or doctor's office for preventative care (all other responses).

We used logit regression models to examine the relationship between our predictor variables and each of our 3 separate questions about the survey participants' use of telehealth: looking up health information using the internet, scheduling an appointment with a health care provider using the internet, and communicating with a provider over email using the internet. In the field of machine learning and statistics, a wide range of computer models can be used for predicting clinical outcomes such as logit regression, decision trees, artificial neural networks, and Bayesian networks. We have selected Logit regression because it is a well-known statistical fitting model that is frequently used for modeling medical problems where it is needed to identify the relation between a binary response variable and a set of independent predictor variables [29]. The models were built using the svyglm function within the survey package in R version 4.2.1 (R Foundation for Statistical Computing) in R Studio 2023.03.0+386 (Posit PBC). The svyglm function includes the ability to account for a weighting variable in the data to facilitate population estimates. The stepAIC function within the MASS package was used to identify the best-fit model for this data.

Results

Overview

Young adults (18-35 years old) made up between 32.1% (n=10,140, unweighted) and 31.26% (n=6902, unweighted) of

weighted survey respondents in 2012 and 2017, respectively, as shown in Table 1, while older adults (56-85 years old) accounted for 32.37% (n=12,662, unweighted) in 2012 and 35.63% (n=11,483, unweighted) in 2018. Women made up between 51.72% (n=13,867, unweighted) and 51.87% (n=19,252, unweighted) of the weighted survey respondents over the years.

Table 1. Demographic information broken down by year; given as count of response and percentage of weighted totals[a].

Parameter	2012	2013	2014	2015	2016	2017	2018
Age, n (%)							
Younger adult	10,140 (32.10)	10,183 (31.92)	10,431 (31.85)	9230 (31.73)	8662 (31.68)	6902 (31.26)	6177 (31.34)
Middle-aged	11,723 (35.53)	11,521 (35.31)	12,065 (34.64)	10,894 (34.23)	10,094 (33.60)	8176 (33.27)	7757 (33.02)
Older adult	12,662 (32.37)	12,853 (32.77)	14,201 (33.51)	13,548 (34.05)	14,272 (34.72)	11,664 (35.47)	11,483 (35.63)
Sex, n (%)							
Male	15,273 (48.13)	15,440 (48.16)	16,398 (48.20)	15,071 (48.20)	14,991 (48.23)	12,096 (48.24)	11,550 (48.28)
Female	19,252 (51.87)	19,117 (51.84)	20,299 (51.80)	18,601 (51.80)	18,037 (51.77)	14,646 (51.76)	13,867 (51.72)
Race, n (%)							
AIAN[b]	349 (0.82)	360 (0.83)	377 (0.81)	392 (0.95)	357 (1.01)	307 (1.20)	295 (1.13)
Asian	2183 (5.35)	2153 (5.59)	2129 (5.74)	1983 (5.94)	1670 (6.07)	1402 (6.37)	1350 (6.41)
Black or African American	5319 (11.91)	5361 (12.03)	5173 (12.26)	4673 (12.32)	3685 (12.31)	2980 (12.42)	2974 (12.38)
Multiple races	659 (1.67)	662 (1.57)	734 (1.61)	699 (1.74)	687 (1.93)	529 (1.99)	563 (2.32)
White	25,939 (80.25)	25,935 (79.98)	28,209 (79.57)	25,831 (79.05)	26,524 (78.68)	21,472 (78.93)	20,173 (77.75)
Region, n (%)							
Northeast	5774 (18.20)	5645 (17.52)	5919 (17.31)	5580 (17.45)	5590 (18.30)	4348 (18.31)	4143 (17.34)
Midwest	7193 (22.72)	7070 (22.68)	7809 (22.99)	7102 (22.42)	7345 (22.17)	6350 (21.81)	5949 (21.98)
South	12,536 (36.43)	12,813 (36.93)	12,896 (37.24)	11,646 (37.12)	11,487 (35.65)	9860 (36.21)	9312 (36.90)
West	9022 (22.65)	9029 (22.88)	10073 (22.46)	9344 (23.01)	8606 (23.88)	6184 (23.67)	6013 (23.78)
Living with spouse or partner, n (%)	14,371 (48.15)	14,199 (48.27)	15,424 (48.08)	14,213 (48.11)	14,160 (48.07)	11,400 (48.66)	11,031 (48.89)
Unweighted sample size	34,525	34,557	36,697	33,672	33,028	26,742	25,417
Weighted sample size	234,920,670	237,394,354	239,688,457	242,500,657	245,142,225	246,657,271	249,455,533

[a]The table shows the unweighted sample size for each variable and the related percentage once the weights are applied to the data.

[b]AIAN: American Indian or Alaska Native.

The percentage of frequent internet users increased from 58.77% (n=18,016, unweighted) in 2012 to 73.15% (n=17,153, unweighted) in 2018. The percentage of email users increased from 64.46% (n=20,038, unweighted) in 2012 to 74.37% (n=17,635, unweighted) in 2018. The number of individuals who reported living with their spouse or partner ranged from 48.07% (n=14,160, unweighted) of weighted survey respondents in 2016 to 48.89% (n=11,031, unweighted) in 2018.

As shown in Figure 1A, older adult respondents tend to look up health information less than the middle-aged or younger adult groups over the years, while the percentage of the respondents looking up health information using the internet grew at similar rates for all groups. As shown in Figure 1B, men tend to look up health information using the internet less than women across all years of the study. Older adult respondents tended to report scheduling an appointment using less than younger adults or middle-aged respondents (Figure 2A). Across the years included in this study, the proportion of younger and middle-aged respondents tended to grow at similar rates. As shown in Figure 2B, men tended to report scheduling an appointment using the internet less often than women. As shown in Figure 3A, a higher percentage of middle-aged adults tended to report communicating with their care provider through email than older adults; a higher percentage of older adults tended to report it than younger adults. As shown in Figure 3B, men tended to report communicating with their care provider through email less often than women.

Figure 1. Percentage of respondents that looked up health information using the internet (A) for each age group and (B) by sex.

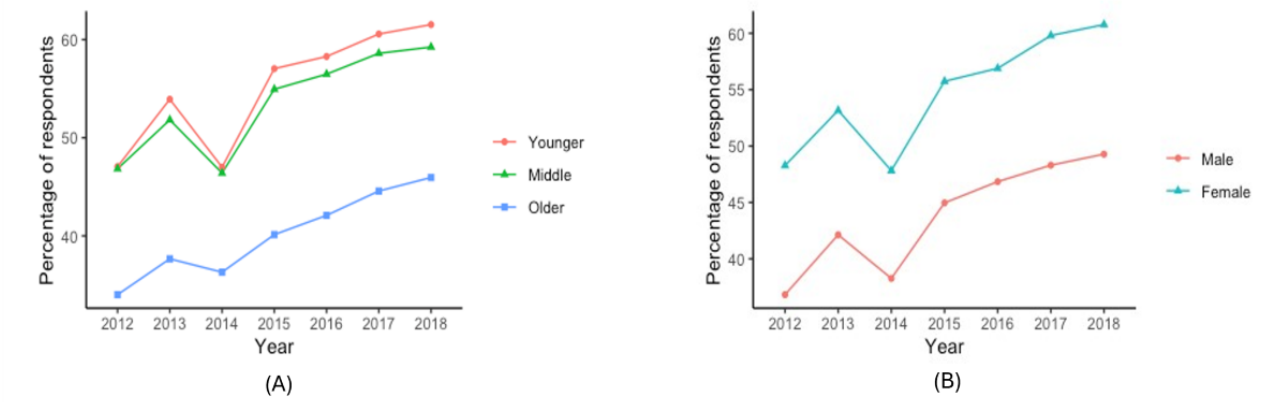


Figure 2. Percentage of respondents that scheduled an appointment using the internet (A) for each age group and (B) by sex.

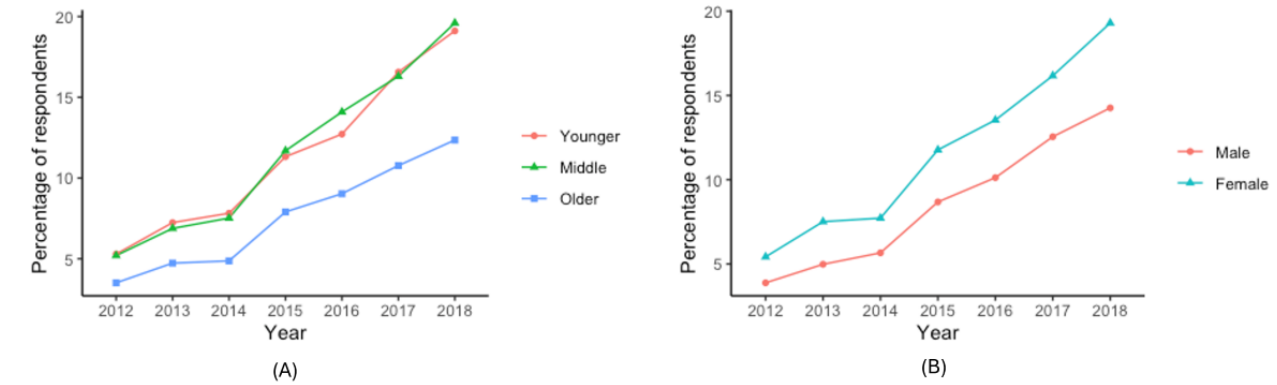
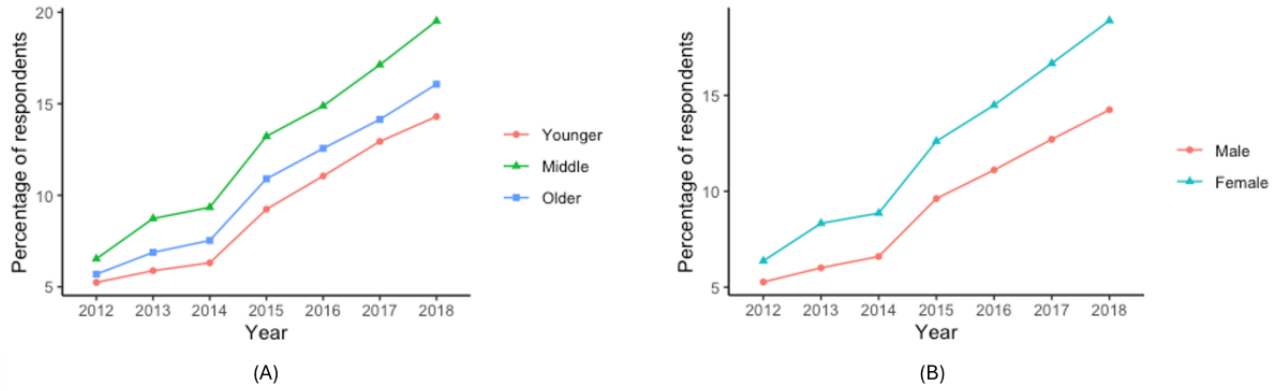


Figure 3. Percentage of respondents that emailed their care provider using the internet (A) for each age group and (B) by sex.



The number of respondents who accessed health information by using the internet increased from 42.76% (n=13,621, unweighted) in 2012 to 55.22% (n=13,677, unweighted) in 2018 (Table 2). In 2018, the percentage of respondents who scheduled an appointment using the internet rose from 4.68% (n=1463, unweighted) in 2012 to 16.86% (n=3962, unweighted). The number of respondents emailing their provider using the internet increased from 5.84% (n=1800, unweighted) in 2012 to 16.65% (n=4176, unweighted) in 2018. The number of respondents seeking medical care who visited a clinic or doctor's office increased from 79.95% (n=27,085, unweighted) in 2012 to 82.03% (n=21,092, unweighted) in 2018. The number of respondents seeking preventive care who visited a clinic or doctor's office increased from 35.96% (n=2706, unweighted) in 2012 to 45.16% (n=2133, unweighted) in 2018. The number of respondents concerned about paying medical bills decreased from 50.24% (n=17,267, unweighted) in 2012 to 43.54% (n=10,637, unweighted) in 2018. The use of alternative therapies to save money increased from 4.07% (n=1454, unweighted) in 2012 to 5.14% (n=1294, unweighted) in 2018.

Table 2. Response variables broken down by year; given as the count of responses and percentage of weighted totals[a].

Parameter	2012	2013	2014	2015	2016	201	2018
Uses internet frequently (>daily), n (%)	18,016 (58.77)	19,699 (63.28)	21,463 (65.04)	19,380 (67.12)	21,013 (69.45)	17,596 (71.05)	17,153 (73.15)
Uses email, n (%)	20,038 (64.46)	21,255 (67.51)	22,697 (68.24)	20,008 (68.84)	21,873 (71.44)	18,198 (72.51)	17,635 (74.37)
Search health information using the internet, n (%)	13,621 (42.76)	15,241 (47.86)	14,783 (43.21)	15,917 (50.55)	16,543 (52.04)	14,097 (54.25)	13,677 (55.22)
Schedule appointment using the internet, n (%)	1463 (4.68)	1973 (6.29)	2099 (6.73)	2974 (10.28)	3542 (11.89)	3543 (14.42)	3962 (16.86)
Email care providers, n (%)	1800 (5.84)	2252 (7.21)	2550 (7.77)	3295 (11.17)	3994 (12.86)	3822 (14.76)	4176 (16.65)
Skip medication to save money, n (%)	2293 (6.31)	1745 (7.72)	1654 (6.93)	1400 (6.07)	1340 (5.78)	1071 (5.96)	1003 (5.71)
Take less medication to save money, n (%)	2418 (6.67)	1875 (8.21)	1733 (7.21)	1519 (6.46)	1390 (6.01)	1118 (6.08)	1074 (5.95)
Request lower-cost medication, n (%)	6320 (18.72)	5186 (24.39)	4918 (21.35)	4259 (19.45)	4460 (19.65)	3456 (18.74)	3436 (19.20)
Uses international medications to save money, n (%)	663 (1.90)	588 (1.57)	554 (1.45)	483 (1.31)	477 (1.62)	397 (1.51)	388 (1.58)
Alternate therapies to save money, n (%)	1454 (4.07)	1559 (4.20)	1587 (4.04)	1358 (3.78)	1411 (4.26)	1161 (4.20)	1294 (5.14)
Cannot afford medications, n (%)	3040 (8.29)	2838 (7.81)	2623 (6.89)	2243 (6.35)	2151 (6.19)	1661 (6.07)	1689 (6.30)
Cannot afford mental care, n (%)	944 (2.51)	798 (2.10)	766 (1.92)	682 (1.87)	686 (1.95)	582 (2.11)	665 (2.65)
Cannot afford dental care, n (%)	4776 (13.23)	4748 (12.85)	4490 (11.60)	3913 (10.90)	3498 (10.24)	2891 (10.77)	2922 (11.26)
Cannot afford vision care, n (%)	2860 (7.80)	2734 (7.33)	2574 (6.47)	2398 (6.40)	1971 (5.64)	1607 (5.81)	1676 (6.33)
Worried about paying medical bills, n (%)	17,237 (50.24)	17,106 (49.76)	16,912 (46.87)	15,136 (45.55)	13,816 (43.72)	11,296 (44.44)	10,637 (43.54)
Goes to a clinic when sick, n (%)							
No where	5660 (16.17)	5325 (15.27)	5033 (13.69)	4518 (13.80)	3941 (13.07)	3167 (12.98)	3137 (13.69)
Somewhere else	1442 (3.88)	1551 (3.90)	1543 (3.97)	1304 (3.59)	1108 (3.27)	947 (3.48)	1047 (4.29)
Goes to clinic	27,085 (79.95)	27,486 (80.83)	29,841 (82.34)	27,587 (82.60)	27,757 (83.66)	22,452 (83.54)	21,092 (82.03)
Visits clinic for preventive care	2706 (35.96)	2478 (33.78)	2607 (35.74)	2495 (39.04)	2640 (44.76)	2191 (46.10)	2133 (45.16)
Unweighted sample size	34,525	34,557	36,697	33,672	33,028	26,742	25,417
Weighted sample size	234,920,670	237,394,354	239,688,457	242,500,657	245,142,225	246,657,271	249,455,533

[a]The table shows the unweighted sample size for each variable and the related percentage once the weights are applied to the data.

Looking Up Health Information Using the Internet

We constructed a model to examine the relationship between the predictor variables and whether or not the survey respondent accessed health information using the internet (Table 3). With each successive year, respondents were more likely to look up health information using the internet (odds ratio [OR] 1.12, 95% CI 1.10-1.15). Middle-aged adults (OR 0.83, 95% CI 0.75-0.93) and older adults (OR 0.65, 95% CI 0.57-0.73) were less likely in comparison to younger adults to look up health information using the internet. Women were more likely to look up health information using the internet in comparison to men (OR 1.69, 95% CI 1.54-1.85). Black or African American respondents were less likely to look up health information in comparison to White respondents (OR 0.77, 95% CI 0.67-0.89). The survey respondents in the South region were less likely to look up health information using the internet compared to those in the

North region (OR 0.73, 95% CI 0.63-0.85). A frequent internet user was more likely to look up health information (OR 3.40, 95% CI 3.00-3.85). Similarly, respondents who used email regularly were more likely to look up health information using the internet than those who reported not using email frequently (OR 3.88, 95% CI 3.41-4.42). Respondents who reported asking their clinicians for lower-cost medications in order to save money were more likely across all years to look up health information (OR 1.55, 95% CI 1.38-1.75). Respondents using alternative therapies in order to save money were almost twice as likely to look up health information using the internet (OR 1.92, 95% CI 1.57-2.36). Respondents who reported going to a clinic or doctor's office when they were sick were more likely to look up health information using the internet than those who did not report going there (OR 1.21, 95% CI 1.08-1.36). Respondents who reported going to a clinic or doctor's office as part of preventative care were also more likely to look up health information using the internet than those who did not seek out preventative care at a clinic or doctor's office (OR 1.18, 95% CI 1.07-1.31).

Table 3. The best-fit model predicting the likelihood that an individual look up health information using the internet.

Parameter	Parameter estimate (SE)	*t* test	*P* value	OR[a] (95% CI)
Intercept	–238.10 (21.92)	–10.86	<.001	N/A[b]
Year	0.12 (0.01)	10.77	<.001	1.12 (1.10-1.15)
Middle-aged	–0.18 (0.05)	–3.35	<.001	0.83 (0.75-0.93)
Older adult	–0.44 (0.06)	–6.81	<.001	0.65 (0.57-0.73)
Female	0.52 (0.05)	11.25	<.001	1.69 (1.54-1.85)
Black or African American	–0.26 (0.07)	–3.62	<.001	0.77 (0.67-0.89)
AIAN[c]	–0.32 (0.24)	–1.33	.18	NS[d]
Asian	–0.28 (0.11)	–2.64	.008	NS
Multiple race	0.00 (0.14)	0.03	.98	NS
Midwest	–0.14 (0.08)	–1.65	.10	NS
South	–0.31 (0.076)	–4.16	<.001	0.73 (0.63-0.85)
West	–0.11 (0.08)	–1.31	.19	NS
Not living with spouse	–0.15 (0.049)	–3.10	.002	NS
Uses internet frequently	1.23 (0.06)	19.29	<.001	3.40 (3.00-3.85)
Uses email	1.36 (0.07)	20.60	<.001	3.88 (3.41-4.42)
Used lower-cost medication to save money	0.44 (0.06)	7.19	<.001	1.55 (1.38-1.75)
Used drugs from other countries to save money	–0.30 (0.15)	–1.95	.05	NS
Used alternate therapies to save money	0.65 (0.10)	6.28	<.001	1.92 (1.57-2.36)
Cannot afford mental care	0.26 (0.11)	2.46	.01	NS
Cannot afford dental care	0.20 (0.07)	2.88	.004	NS
Cannot afford eye care	0.12 (0.08)	1.44	.15	NS
Worried about paying for medical bills	0.08 (0.05)	1.70	.09	NS
Goes somewhere other than a clinic or doctor's office when sick	0.13 (0.07)	1.82	.07	NS
Goes to a clinic or doctor's office when sick	0.19 (0.05)	3.36	<.001	1.21 (1.08-1.36)
Goes to a clinic or doctor's office for preventative care	0.17 (0.05)	3.39	<.001	1.18 (1.07-1.31)

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]AIAN: American Indian or Alaska Native.

[d]NS: parameter was not significant at α=.001.

Scheduling an Appointment Using the Internet

We constructed a model to examine the relationship between the predictor variables and whether or not the survey respondent scheduled an appointment using the internet (Table 4). With each successive year, respondents were more likely to schedule an appointment using the internet (OR 1.30, 95% CI 1.26-1.34). Older adults were less likely in comparison to younger adults to schedule an appointment using the internet (OR 0.63, 95% CI 0.51-0.78). Women were more likely than men to schedule an appointment using the internet (OR 1.61, 95% CI 1.40-1.85). Respondents who use the internet frequently were more likely to schedule an appointment than those not using the internet frequently (OR 2.75, 95% CI 2.02-3.74). Survey respondents using email were more likely than respondents not using email to schedule an appointment using the internet (OR 3.92, 95% CI 2.78-5.53). Respondents who used alternative therapies in order to save money were more likely to schedule an appointment using the internet than those who did not report using alternative therapies to save money (OR 1.70, 95% CI 1.34-2.16). Respondents who were worried about paying medical bills in the last 12 months were less likely than those who did not report being worried about paying medical bills to schedule an appointment using the internet (OR 0.77, 95% CI 0.66-0.89). Respondents going to a clinic or doctor's office for preventative care were more likely to schedule an appointment using the internet than those who did not report going to a clinic or doctor's office for preventative care (OR 1.55, 95% CI 1.33-1.82).

Table 4. The best-fit model predicting the likelihood that an individual schedules a health care appointment using the internet.

Parameter	Parameter estimate (SE)	*t* test	*P* value	OR[a] (95% CI)
Intercept	–537.38 (32.31)	–16.63	<.001	N/A[b]
Year	0.26 (0.02)	16.49	<.001	1.30 (1.26-1.34)
Middle-aged	–0.19 (0.08)	–2.37	.02	NS[c]
Older adult	–0.46 (0.11)	–4.22	<.001	0.63 (0.51-0.78)
Female	0.48 (0.07)	6.63	<.001	1.61 (1.40-1.85)
Black or African American	–0.08 (0.11)	–0.66	.51	NS
AIAN[d]	–0.71 (0.53)	–1.34	.18	NS
Asian	0.40 (0.14)	2.88	.004	NS
Multiple race	0.20 (0.23)	–0.89	.37	NS
Midwest	–0.37(0.12)	–3.01	.003	NS
South	–0.10 (0.11)	–0.92	.36	NS
West	0.03 (0.11)	0.26	.79	NS
Not living with spouse or partner	–0.08 (0.07)	–1.10	.28	NS
Uses internet frequently (>1 per day)	1.01 (0.16)	6.46	<.001	2.75 (2.02-3.74)
Uses email	1.37 (0.18)	7.80	<.001	3.92 (2.78-5.53)
Used lower-cost medication to save money	0.26 (0.09)	2.96	.003	NS
Used alternate therapies to save money	0.53 (0.12)	4.39	<.001	1.70 (1.34-2.16)
Cannot afford prescription medicine	–0.23 (0.14)	–1.67	.09	NS
Cannot afford mental care	0.37 (0.17)	2.21	.03	NS
Cannot afford dental care	–0.28 (0.11)	–2.40	.02	NS
Worried about paying for medical bills	–0.26 (0.08)	–3.46	<.001	0.77 (0.66-0.89)
Goes somewhere other than a clinic or doctor's office when sick	–0.07 (0.12)	–0.55	.58	NS
Goes to a clinic or doctor's office when sick	0.18 (0.08)	2.16	.03	NS
Goes to a clinic or doctor's office for preventative care	0.44 (0.08)	5.50	<.001	1.55 (1.33-1.82)

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]NS: parameter was not significant at α=.001.

[d]AIAN: American Indian or Alaska Native.

Communicating With a Care Provider Through Email Using the Internet

We constructed a model to examine the relationship between the predictor variables and whether or not the survey respondent communicated with a care provider through email using the internet (Table 5). With each successive year, respondents were more likely to communicate with a care provider through email using the internet (OR 1.24, 95% CI 1.20-1.28). Women were more likely than men to communicate with a care provider through email using the internet (OR 1.48, 95% CI 1.28-1.71). Respondents who use the internet frequently were more likely to communicate with a care provider through email than those not using the internet frequently (OR 2.08, 95% CI 1.56-2.76). Survey respondents using email were more likely than respondents not using email to communicate with a care provider through email (OR 5.56, 95% CI 3.84-8.05). Respondents who could not afford mental care in the last 12 months were more likely to communicate with a care provider through email using the internet than those who could afford mental care in the last 12 months (OR 1.92, 95% CI 1.39-2.66). Respondents who reported going to a clinic or doctor's office when they were sick were more likely to communicate with a care provider through email using the internet than those who did not report going there when sick (OR 1.32, 95% CI 1.12-1.55). Respondents who reported going to a clinic or doctor's office as part of preventative care were also more likely to communicate with a care provider through email using the internet than those who did not seek out preventative care at a clinic or doctor's office (OR 1.57, 95% CI 1.34-1.85).

Table 5. The best-fit model predicting the likelihood that an individual communicated with a care provider through email using the internet.

Parameter	Parameter estimate (SE)	*t* test	*P* value	OR[a] (95% CI)
Intercept	–442.69 (32.92)	–13.45	<.001	N/A[b]
Year	0.22 (0.02)	13.29	<.001	1.24 (1.20-1.28)
Middle-aged	0.16 (0.08)	1.88	.06	NS[c]
Older adult	0.18 (0.11)	1.71	.09	NS
Female	0.39 (0.07)	5.34	<.001	1.48 (1.28-1.71)
Black or African American	0.02 (0.13)	0.14	.89	NS
AIAN[d]	–0.37 (0.42)	–0.90	.37	NS
Asian	0.37 (0.15)	2.50	.01	NS
Multiple race	–0.44 (0.24)	–1.80	.07	NS
Midwest	–0.09 (0.13)	–0.72	.47	NS
South	–0.09 (0.12)	–0.79	.43	NS
West	0.32 (0.12)	2.71	.007	NS
Not living with spouse or partner	–0.19 (0.07)	–2.55	.01	NS
Uses internet frequently (>1 per day)	0.73 (0.14)	5.04	<.001	2.08 (1.56-2.76)
Uses email	1.72 (0.19)	9.10	<.001	5.56 (3.84-8.05)
Skipped medication to save money	0.31 (0.20)	1.55	.12	NS
Took less medication to save money	–0.30 (0.21)	–1.43	.15	NS
Used lower-cost medication	0.30 (0.09)	3.18	.002	NS
Used alternate therapies	0.41 (0.13)	3.11	.002	NS
Cannot afford medication	–0.24 (0.16)	–1.50	.13	NS
Cannot afford mental care	0.65 (0.16)	3.98	<.001	1.92 (1.39-2.66)
Cannot afford dental care	–0.30 (0.13)	–2.26	.02	0.74 (0.57-0.96)
Cannot afford eye care	–0.19 (0.17)	–1.14	.26	NS
Worried about a medical bill	–0.18 (0.08)	–2.35	.02	NS
Goes somewhere other than a clinic or doctor’s office when sick	0.10 (0.12)	0.85	.39	NS
Goes to a clinic or doctor’s office when sick	0.28 (0.08)	3.32	<.001	1.32 (1.12-1.55)
Goes to a clinic or doctor’s office for preventative care	0.45 (0.08)	5.50	<.001	1.57 (1.34-1.85)

[a]OR: odds ratio.

[b]N/A: not applicable.

[c]NS: parameter was not significant at α=.001.

[d]AIAN: American Indian or Alaska Native.

Discussion

Principal Findings

This study used NHIS data, which is weighted to represent national characteristics, to examine the use of technology for performing some health care–related tasks over time: looking up health information using the internet, scheduling an appointment using the internet, and communicating with a care provider over email using the internet. Across all our models, we found that middle-aged and older adult respondents were significantly less likely to use technology to look up health information using the internet or schedule an appointment using the internet versus younger individuals. It has been shown that some older adults may need additional special usability requirements due to their inexperience in the use of technology [21], yet this population is a growing user of technology [30]. This also reaffirmed the results found in the existing literature [26]. Across all of the variables we investigated, we found that the rates of looking up health information using the internet, scheduling an appointment using the internet, and communicating with a care provider over email using the internet increased substantially across the study period. This demonstrates that there is an increasing use and need for these

services to support larger populations of users including devices and abilities to engage with technology.

Specifically, our analysis found that middle-aged adults were found to have a higher OR than older adults (0.83 vs 0.65) for looking up health information using the internet. We also found that there were differences in age groups for using technology to perform health care–related tasks. In terms of searching health information and scheduling appointments using the internet, we found differences between men and women, with women being significantly more likely than men to look up health information using the internet, schedule an appointment using the internet, and communicate with a care provider through email using the internet. There are conflicting results in the literature related to sex differences in the use of the internet and technology associated with health-related tasks. Newhouse et al [31] found that men were more likely to use email communication for health care than women; however, Baumann et al [32] found that women were more likely to use the internet and technology for health-related tasks such as scientific literature review, communicating with their physician or medical team, and interpreting the diagnostic test results and medications used in treatment. These sex differences may be related to the social construction and perceptions of technology [33]. Future research needs to explore why these trends are occurring and what factors are associated with the differences between men and women especially since there are conflicting results in the literature about use but also about the factors associated with the differences in use.

The impact of costs was inconsistent across the different models in our analysis. With respect to looking up health information using the internet, there was a significant association with using lower-cost medications and alternative therapies to save money. For individuals scheduling a health care appointment using the internet, respondents who indicated using alternative therapies to save money were more likely to schedule appointments using the internet, and surprisingly those who indicated being worried about paying for medical bills were less likely to schedule appointments using the internet. Individuals using email to communicate with a care provider had two different cost-related variables that were significant. Respondents who reported not being able to afford mental care were more likely to email, while those who indicated not being able to afford dental care were less likely to email. In some ways, the included cost-related variables did not seem to indicate a consistent response around cost. This reinforces the notion that determining the extent to which the cost of care is a barrier is challenging and unclear [34]. By examining the results across several cost-of-care–related variables and different elements of health care, it appears that there may be other barriers whose interaction with cost impacts the effect of cost as a barrier. As suggested by Clarke et al [34], and supported by our results, telehealth research should continue to investigate the ways in which different types of costs are intertwined with the use of telehealth; especially as it relates to different avenues of care.

Going to a clinic or doctor's office for preventative care was associated with a greater likelihood of looking up health information using the internet, schedule a health care appointment using the internet, and communicate with a care provider through email using the internet. Going to a clinic or doctor's office when sick was associated with using technology to look up health information and communicating with a care provider through email. This suggests that there is a strong correlation between respondents' collaboration in their personal health and the likelihood that they would use telehealth services to meet these needs. In fact, Sawesi et al [35] found that information technology platforms can enhance patient engagement and improve health outcomes. It is crucial that telehealth research investigate ways in which telehealth can be used to either support individuals in developing an engagement in their own health or identify ways to encourage users to develop an engagement in their own health.

Limitations and Future Work

There are several factors that limit the generalizability of our analyses. One notable limitation of this research is that it relies on self-reported survey data and may not fully capture specific perspectives and opinions about the "why" for some of the activities reported in the NHIS. Yet, the NHIS data collected and maintained by the US CDC is widely used [36-38] and the data, the data design, and the imputation for national representation are widely documented [39,40]. Future research should expand on these results to try to identify if these trends continue or what factors may be driving the differences identified over time.

This research uses responses from specific questions from 2012 to 2018, as those were the only consecutive years where these questions were consistent. As a result, being able to project current and future use of technology for telehealth is limited. Additionally, the COVID-19 pandemic had a major impact on health care as it presented a unique situation that resulted in overcoming many of the traditional barriers to telehealth adoption [41]. COVID-19 also emphasized the use cases of a variety of different technological tools that can be used in telehealth such as chatbots [42]. As our data was limited to data collected before the COVID-19 pandemic, it is important that future work examine the impact that the pandemic has had on the use of telehealth services and if these effects have been sustained in the following years. Future research should also investigate the ways in which the pandemic might have impacted or altered the trends and barriers examined in this research. For example, it may be interesting to evaluate if the exposure and use of these tools during the pandemic have had lasting impacts on the use of the internet tools. To do this, we need similar data from 2020-2023 or later from the NHIS on the use of telehealth services to be able to apply time series modeling techniques to evaluate specific trends in how using telehealth services have changed over time. Interestingly, some initial work has suggested that there is a decline in the use of telemedicine between 2021 and 2022 in the United States, [43] and future work should evaluate if that trend continues or if the trend prior to the COVID-19 pandemic continues.

In order to develop a fundamental understanding of the important barriers, this research did not examine the complex relationships between demographic and response variables. Future research should focus on understanding how these interactions (eg, older women vs middle-aged women vs

younger women) might be correlated with the respondents' use of telehealth. Our models also did not investigate telehealth use for specific socioeconomic levels. Instead, it includes questions regarding a respondent's ability to afford different health care services, which is different from financial security or socioeconomic conditions [44].

Conclusions

As telehealth is increasingly becoming an important component of health care services, it is important to focus on different aspects of telehealth to determine the demographic characteristics, behaviors, or opinions that may predict or influence groups that are likely to face a barrier to using telehealth services. This study used NHIS data to examine the use of technology for performing some health care–related tasks over time: looking up health information using the internet, scheduling an appointment using the internet, and communicating with a care provider over email using the internet. From this analysis, we have found some potential barriers that may impact a user's ability to access telehealth services, as well as differences in the use of these tools for different groups. Understanding those who are using the internet for health care–related activities and the barriers that they may face is important for the design and implementation of these systems to be as effective as possible. Systems and technology designers, as well as health care providers, should be aware of these differences and the impacts they may have on engaging individuals in their health care.

Conflicts of Interest

None declared.

References

1. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014;11(2):1279-1298 [FREE Full text] [doi: 10.3390/ijerph110201279] [Medline: 24464237]
2. Hu Y. Health communication research in the digital age: a systematic review. J Commun Healthcare 2015;8(4):260-288. [doi: 10.1080/17538068.2015.1107308]
3. Doarn CR, Pruitt S, Jacobs J, Harris Y, Bott DM, Riley W, et al. Federal efforts to define and advance telehealth--a work in progress. Telemed J E Health 2014 May;20(5):409-418. [doi: 10.1089/tmj.2013.0336] [Medline: 24502793]
4. Maheu M, Whitten P, Allen A. E-Health, telehealth, and telemedicine. J Healthcare Qual 2002;24(4):52. [doi: 10.1111/j.1945-1474.2002.tb00449.x]
5. Tuckson RV, Edmunds M, Hodgkins ML. Telehealth. N Engl J Med 2017;377(16):1585-1592. [doi: 10.1056/NEJMsr1503323] [Medline: 29045204]
6. Gogia S. Rationale, history, and basics of telehealth. Fundam Telemed Telehealth 2020:11-34. [doi: 10.1016/b978-0-12-814309-4.00002-1]
7. Hoffman DA. Increasing access to care: telehealth during COVID-19. J Law Biosci 2020;7(1):lsaa043 [FREE Full text] [doi: 10.1093/jlb/lsaa043] [Medline: 32843985]
8. Mahtta D, Daher M, Lee MT, Sayani S, Shishehbor M, Virani SS. Promise and perils of telehealth in the current era. Curr Cardiol Rep 2021;23(9):115 [FREE Full text] [doi: 10.1007/s11886-021-01544-w] [Medline: 34269884]
9. Curfman AL, Hackell JM, Herendeen NE, Alexander JJ, Marcin JP, Moskowitz WB, et al. Telehealth: improving access to and quality of pediatric health care. Pediatrics 2021;148(3):e2021053129 [FREE Full text] [doi: 10.1542/peds.2021-053129] [Medline: 34462339]
10. Finley BA, Shea KD. Telehealth: disrupting time for health care quantity and quality. Nurs Adm Q 2019;43(3):256-262. [doi: 10.1097/NAQ.0000000000000357] [Medline: 31162344]
11. Uscher-Pines L, Rudin RS. Lessons learned from the MAVEN project pilot: using physician volunteers to increase access to care via telehealth. Rand Health Q 2020;9(1):9 [FREE Full text] [Medline: 32742751]
12. Alford-Teaster J, Wang F, Tosteson ANA, Onega T. Incorporating broadband durability in measuring geographic access to health care in the era of telehealth: a case example of the 2-step virtual catchment area (2SVCA) method. J Am Med Inform Assoc 2021;28(11):2526-2530 [FREE Full text] [doi: 10.1093/jamia/ocab149] [Medline: 34414437]
13. Boothe E, Kaplan J. Using telemedicine in Mississippi to improve patient access to genetic services. J Genet Couns 2018;27(2):320-322. [doi: 10.1007/s10897-017-0192-6] [Medline: 29264692]
14. Chen YH, Lin YH, Hung CS, Huang CC, Yeih DF, Chuang PY, et al. Clinical outcome and cost-effectiveness of a synchronous telehealth service for seniors and nonseniors with cardiovascular diseases: quasi-experimental study. J Med Internet Res 2013;15(4):e87 [FREE Full text] [doi: 10.2196/jmir.2091] [Medline: 23615318]
15. Langarizadeh M, Tabatabaei M, Tavakol K, Naghipour M, Rostami A, Moghbeli F. Telemental health care, an effective alternative to conventional mental care: a systematic review. Acta Inform Med 2017;25(4):240-246 [FREE Full text] [doi: 10.5455/aim.2017.25.240-246] [Medline: 29284913]
16. Brown SH, Griffith ML, Kripalani S, Horst SN. Association of health literacy and area deprivation with initiation and completion of telehealth visits in adult medicine clinics across a large health care system. JAMA Netw Open 2022;5(7):e2223571 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.23571] [Medline: 35877125]

17. Irfan A, Lever JM, Fouad MN, Sleckman BP, Smith H, Chu DI, et al. Does health literacy impact technological comfort in cancer patients? Am J Surg 2022;223(4):722-728 [FREE Full text] [doi: 10.1016/j.amjsurg.2021.08.006] [Medline: 34384588]
18. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
19. Sullivan-Baca E, Babicz MA, Choudhury TK, Miller BI. The relationship between health literacy and comfort with teleneuropsychology in a veteran sample. Arch Clin Neuropsychol 2022;37(2):292-301 [FREE Full text] [doi: 10.1093/arclin/acab079] [Medline: 34599332]
20. Demiris G, Charness N, Krupinski E, Ben-Arieh D, Washington K, Wu J, et al. The role of human factors in telehealth. Telemed e-Health 2010;16(4):446-453. [doi: 10.1089/tmj.2009.0114] [Medline: 20420540]
21. Dhillon JS, Ramos C, Wunsche BC, Lutteroth C. Designing a web-based telehealth system for elderly people: an interview study in New Zealand. : IEEE; 2011 Presented at: 24th International Symposium on Computer-Based Medical Systems (CBMS); June 27, 2011; Bristol, UK. [doi: 10.1109/cbms.2011.5999157]
22. Zhang T, Mosier J, Subbian V. Identifying barriers to and opportunities for telehealth implementation amidst the COVID-19 pandemic by using a human factors approach: a leap into the future of health care delivery? JMIR Hum Factors 2021;8(2):e24860 [FREE Full text] [doi: 10.2196/24860] [Medline: 33779566]
23. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]
24. Lin CCC, Dievler A, Robbins C, Sripipatana A, Quinn M, Nair S. Telehealth in health centers: key adoption factors, barriers, and opportunities. Health Aff (Millwood) 2018;37(12):1967-1974. [doi: 10.1377/hlthaff.2018.05125] [Medline: 30633683]
25. Weber E, Miller SJ, Astha V, Janevic T, Benn E. Characteristics of telehealth users in NYC for COVID-related care during the coronavirus pandemic. J Am Med Inform Assoc 2020;27(12):1949-1954 [FREE Full text] [doi: 10.1093/jamia/ocaa216] [Medline: 32866249]
26. Hung L, Lyons JG, Wu C. Health information technology use among older adults in the United States, 2009-2018. Curr Med Res Opin 2020;36(5):789-797. [doi: 10.1080/03007995.2020.1734782] [Medline: 32096650]
27. Haughton B, Stang J. Population risk factors and trends in health care and public policy. J Acad Nutr Diet 2012;112(3 Suppl):S35-S46. [doi: 10.1016/j.jand.2011.12.011] [Medline: 22709860]
28. About the National Health Interview Survey. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/nchs/nhis/about_nhis.htm [accessed 2024-10-18]
29. Ayer T, Chhatwal J, Alagoz O, Kahn CE, Woods RW, Burnside ES. Informatics in radiology: comparison of logistic regression and artificial neural network models in breast cancer risk estimation. Radiographics 2010;30(1):13-22 [FREE Full text] [doi: 10.1148/rg.301095057] [Medline: 19901087]
30. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]
31. Newhouse N, Lupiáñez-Villanueva F, Codagnone C, Atherton H. Patient use of email for health care communication purposes across 14 European countries: an analysis of users according to demographic and health-related factors. J Med Internet Res 2015;17(3):e58 [FREE Full text] [doi: 10.2196/jmir.3700] [Medline: 25798912]
32. Baumann E, Czerwinski F, Reifegerste D. Gender-specific determinants and patterns of online health information seeking: results from a representative German health survey. J Med Internet Res 2017;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
33. Noone JH, Stephens C. Men, masculine identities, and health care utilisation. Sociol Health Illn 2008;30(5):711-725 [FREE Full text] [doi: 10.1111/j.1467-9566.2008.01095.x] [Medline: 18564976]
34. Clarke M, Fursse J, Brown-Connolly NE, Sharma U, Jones R. Evaluation of the National Health Service (NHS) direct pilot telehealth program: cost-effectiveness analysis. Telemed e-Health 2018;24(1):67-76. [doi: 10.1089/tmj.2016.0280] [Medline: 28723244]
35. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
36. Peprah E, Gyamfi J, Lee JT, Islam F, Opeyemi J, Tampubolon S, et al. Analysis of the 2007-2018 National Health Interview Survey (NHIS): examining neurological complications among children with sickle cell disease in the United States. Int J Environ Res Public Health 2023;20(12):6137 [FREE Full text] [doi: 10.3390/ijerph20126137] [Medline: 37372724]
37. Blumberg SJ, Parker JD, Moyer BC. National Health Interview Survey, COVID-19, and online data collection platforms: adaptations, tradeoffs, and new directions. Am J Public Health 2021;111(12):2167-2175. [doi: 10.2105/AJPH.2021.306516] [Medline: 34878857]
38. Gyamfi J, Tampubolon S, Lee JT, Islam F, Ojo T, Opeyemi J, et al. Characterisation of medical conditions of children with sickle cell disease in the USA: findings from the 2007-2018 National Health Interview Survey (NHIS). BMJ Open 2023;13(2):e069075 [FREE Full text] [doi: 10.1136/bmjopen-2022-069075] [Medline: 36854589]

39. Blewett LA, Dahlen HM, Spencer D, Rivera Drew JA, Lukanen E. Changes to the design of the National Health Interview Survey to support enhanced monitoring of health reform impacts at the state level. Am J Public Health 2016;106(11):1961-1966 [FREE Full text] [doi: 10.2105/AJPH.2016.303430] [Medline: 27631739]
40. Kindratt TB. National Health Interview Survey. Mavs Open Press. 2022. URL: https://uta.pressbooks.pub/bigdataforepidemiology/chapter/chapter6-nhis/ [accessed 2024-08-09]
41. Hoffman DA. Increasing access to care: telehealth during COVID-19. J Law Biosci 2020;7(1):lsaa043 [FREE Full text] [doi: 10.1093/jlb/lsaa043] [Medline: 32843985]
42. Espinoza J, Crown K, Kulkarni O. A guide to chatbots for COVID-19 screening at pediatric health care facilities. JMIR Public Health Surveill 2020;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]
43. Lucas JW, Wang X. Declines in telemedicine use among adults: United States, 2021 and 2022. Natl Health Stat Report 2024(205):1-11 [FREE Full text] [Medline: 38912919]
44. Williams C, Shang D. Telehealth usage among low-income racial and ethnic minority populations during the COVID-19 pandemic: retrospective observational study. J Med Internet Res 2023;25:e43604 [FREE Full text] [doi: 10.2196/43604] [Medline: 37171848]

Abbreviations

CDC: Centers for Disease Control and Prevention
NHIS: National Health Interview Survey
OR: odds ratio

Edited by K Adapa; submitted 13.03.24; peer-reviewed by S Sawesi, EP Greenup; comments to author 11.07.24; revised version received 09.08.24; accepted 23.08.24; published 26.02.25.

Please cite as:
Junkins Z, Zahan N, Neyens D
Examining Individuals' Use of the Internet for Health Care Activities Over Time: Results from the US National Health Interview Survey
JMIR Hum Factors 2025;12:e58362
URL: https://humanfactors.jmir.org/2025/1/e58362
doi:10.2196/58362
PMID:

Original Paper

Development and Systematic Evaluation of a Progressive Web Application for Women With Cardiac Pain: Usability Study

Monica Parry[1], RN-EC, PhD; Tony Huang[1], BScH; Hance Clarke[2,3], MD, PhD, FRCPC; Ann Kristin Bjørnnes[4], RN, PhD; Paula Harvey[3,5], BMBS, PhD, FRACP; Laura Parente[6], BDES; Colleen Norris[7], RN, PhD; Louise Pilote[8], MD, PhD, FRCPC; Jennifer Price[5], RN, PhD; Jennifer N Stinson[1,9], RN-EC, PhD; Arland O'Hara[1], BSc; Madusha Fernando[1], BSc; Judy Watt-Watson[1], RN, PhD; Nicole Nickerson[10]; Vincenza Spiteri DeBonis[11], MES; Donna Hart[11], BA; Christine Faubert[11]

[1]Lawrence Bloomberg Faculty of Nursing, University of Toronto, Toronto, ON, Canada
[2]Pain Research Unit, University Health Network, Toronto, ON, Canada
[3]University of Toronto, Toronto, ON, Canada
[4]Department of Nursing and Health Promotion, Oslo Metropolitan University, Oslo, Norway
[5]Women's College Hospital, Toronto, ON, Canada
[6]Healthcare Human Factors, University Health Network, Toronto, ON, Canada
[7]Faculty of Nursing, University of Alberta, Edmonton, AB, Canada
[8]Department of Medicine, McGill University, Research Institute of the McGill University Health Centre, Montreal, QC, Canada
[9]Research Institute, The Hospital for Sick Children, Toronto, ON, Canada
[10]Patient Advisor, LaHave, NS, Canada
[11]Patient Advisor, Toronto, ON, Canada

Corresponding Author:
Monica Parry, RN-EC, PhD
Lawrence Bloomberg Faculty of Nursing
University of Toronto
Office 276
155 College Street
Toronto, ON, M5T 1P8
Canada
Phone: 1 416 946 3561
Fax: 1 416 978 8222
Email: monica.parry@utoronto.ca

Abstract

Background: Cardiac pain has been widely considered to be the primary indicator of coronary artery disease. The presentation of cardiac pain and associated symptoms vary in women, making it challenging to interpret as cardiac, possibly cardiac, or noncardiac. Women prefer to consult with family and friends instead of seeking immediate medical care.

Objective: This study aimed to assess the user performance (ie, ease of use, efficiency, and errors) and user satisfaction (System Usability Scale; SUS) of a progressive web application for women with cardiac pain.

Methods: Following ethics approval, a purposive sample of women aged >18 years with cardiac pain or associated symptoms lasting >3 months and able to speak and read English was recruited to participate in 2 iterative usability testing cycles. The first cycle assessed the performance of and satisfaction with *at heart* using a web application, and the second cycle assessed the performance of and satisfaction with *at heart* across various Android and iOS devices. In total, 2 investigators recorded user comments and documented problems. At the end of the testing session, the participants completed the SUS and 4 semistructured interview questions.

Results: In total, 10 eligible women participated in usability testing from March 31, 2020, to April 17, 2020 (cycle 1), and from November 17, 2020, to November 30, 2020 (cycle 2). Women across usability testing cycles had a mean age of 55.6 (SD 7.3) years, and most (9/10, 90%) were well educated. In total, 50% (5/10) were employed full or part time, and 60% (6/10) earned >CAD $70,000 (US $48,881.80) annually. Participants across 2 testing cycles reported the overall usability of the *at heart*

progressive web application as highly acceptable (mean SUS score 81.75, SD 10.41). In total, 90% (9/10) of participants rated the user-friendliness of *at heart* as good or excellent. All participants (10/10, 100%) thought *at heart* was easy to use and efficient. Only 2 testing errors were noted as high priority; these were low contrast or small font and clarification that the chatbot was not a real person. User satisfaction was assessed using themes that emerged from the debrief and 4 semistructured interview questions; *at heart* was engaging, comprehensive, understandable, credible, relevant, affirming, personalized, and innovative.

Conclusions: This study provides initial support for the *at heart* progressive web application for women living with cardiac pain and symptoms. Ongoing evaluations in phases 3 and 4 should aim to examine the feasibility and acceptability of and the extent of engagement with the *at heart* core feature set: Heart Check, Wellness Check, and the library. In addition to assessing effectiveness in the phase-4 effectiveness-implementation hybrid trial (type I), describing and better understanding the context for implementation (eg, race and ethnicity and geography) will be necessary.

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-033092

(JMIR Hum Factors 2025;12:e57583) doi:10.2196/57583

KEYWORDS

digital health; chatbot; women; cardiac pain; usability testing; self-management; artificial intelligence; AI

Introduction

Background

Cardiovascular diseases constitute the leading cause of mortality worldwide, exerting a substantial economic burden on the health care system [1-4]. As the most prevalent form of cardiovascular disease, coronary artery disease (CAD) is estimated to be associated with 8.93 million deaths annually worldwide across all ages [1]. CAD is a complex condition that varies in clinical presentation across sexes, with both obstructive (macrovascular) and nonobstructive (microvascular) CAD being associated with cardiac pain and other symptoms [5]. Although cardiac pain has been widely considered to be the primary indicator of CAD [6], cardiac pain and associated symptoms reported by women with CAD differ markedly from those reported by men [7,8]. The presentation of cardiac pain and associated symptoms may vary in frequency, pattern, and distribution in women [8], making it challenging to interpret as cardiac specific [9]. Similarly, women that undergo a percutaneous coronary intervention or cardiac surgery report higher prevalence of persistent pain of moderate to severe intensity after treatment as compared to men [10-12]. This cardiac pain is often described by women as sharp and burning and may present with varying extent of dyspnea, fatigue, anxiety, and discomfort that may radiate to the jaws, shoulders, back, and arms [6,8,13]. The manifestations of cardiac pain and associated symptoms contribute to substantial morbidity and impairments in health-related quality of life (HRQoL) in women [14,15].

The complex presentation of CAD in women poses a challenge for timely recognition and management of symptoms. Recent data have shown that women often delay seeking medical care when experiencing acute cardiac pain or associated symptoms, with time between symptom onset and emergency department arrival being 85 to 320 minutes [16]. Women report experiencing difficulties in interpreting, understanding, and attributing their cardiac pain or symptoms to CAD, preferring to consult with family and friends instead of seeking immediate medical care [17]. Moreover, women often hesitate to seek medical care for cardiac pain or associated symptoms as a result of their gendered roles as caregivers [17,18]. Women describe having gendered roles that they are unable to delegate, such as providing care for dependent family members [17,18]. The cumulative effect of symptom underrecognition and hesitancy in women leads to delayed care-seeking behaviors and an increased risk of major adverse cardiac events or mortality as compared to men [6]. As such, it is imperative to promote proper recognition, assessment, and management of symptoms in women to improve health outcomes and HRQoL.

Self-management programs are designed to engage users as active participants in the management of their conditions; they are key predictors of successful behavior change [19,20]. These interventions generally use educational strategies designed to assist users in achieving optimal knowledge, understanding beliefs, and skills, as well as providing meaningful social supports [21]. Digital health–based self-management programs have been developed and effectively used to help women manage weight [22-25], increase physical activity [26], monitor for perinatal depression, and assist with postpartum smoking cessation [27]. Many women describe digital health interventions as being novel and supportive [22] and effective in motivating healthy behaviors, reducing symptoms [28], and improving HRQoL [23]. However, there is a lack of evidence-informed digital health self-management programs specifically for women with CAD living with cardiac pain and associated symptoms [7], demonstrating a clear need for a digital self-management program.

At heart (formerly HEARTPA♀N) [29], a self-management progressive web application, was developed for women with CAD using a sequential phased approach recommended by the Medical Research Council (MRC) [30-32]. In phase 1, an integrated mixed methods systematic review was conducted to evaluate the current evidence related to the self-management of cardiac pain and associated symptoms (eg, dyspnea and fatigue) in women [7,29]. The results of the review suggested that self-management interventions could reduce cardiac pain and associated symptoms if they targeted a greater proportion of women (standardized mean difference [SMD]=−0.01; SE 0.003; *P*=.02), goal setting (SMD=−0.26, 95% CI −0.49 to −0.03), and collaboration or support from a health care provider (SMD=−0.57, 95% CI −1.00 to −0.14) [33]. The review also identified a lack of self-management interventions targeted specifically for cardiac pain and associated symptoms in women

[33]. In phase 2A, the content and core feature set, chatbot, and symptom triage algorithms were co-designed with health care professionals and women with CAD [29]. In phase 2B, the usability of *at heart* was evaluated to ensure that the platform was intuitive and acceptable for women with cardiac pain, and it is the focus of this paper. In phase 3, a process and preliminary efficacy evaluation using a 2-group parallel pilot randomized controlled trial (RCT) will be undertaken, and then a phase-4 effectiveness-implementation hybrid trial is planned (Figure 1).

Figure 1. Phases of the *at heart* progressive web application development.

Objectives

The objectives of this study were to assess the user performance (ie, ease of use, efficiency, and errors) and user satisfaction (interviews and System Usability Scale [SUS]) of a progressive web application for women with cardiac pain.

Methods

Participant Selection

Following ethics approval, a purposive sample of women was recruited from (1) an ambulatory care hospital focused on women's health, (2) an adult tertiary care transitional pain clinic; (3) the Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease registry; (4) the CorHealth Ontario Cardiac Registry; (5) established strategies through the Canadian Pain Coalition; (6) the Ontario Women's Health Network's listserve, which reaches >1900 women and community organizations in Ontario, Canada; and (7) social media. The Ontario Women's Health Network regularly brought women who lived in rural and remote areas together, including women who were disabled, women of color, women with a low income, and Indigenous and older women. Women were eligible to participate in the usability testing if they (1) were aged >18 years, (2) had been diagnosed with obstructive or nonobstructive CAD pain or pain after a percutaneous coronary intervention or cardiac surgery lasting >3 months, (3) were able to speak and read English, and (4) had not previously used or tested the *at heart* progressive web application. Women were excluded from the study if they had (1) severe cognitive impairment, assessed using the Six-Item Screener administered via telephone; or (2) a major comorbid medical or psychiatric condition that would preclude their ability to participate in the usability testing. The Six-Item Screener is a brief instrument for identifying individuals with cognitive impairment, and its diagnostic properties are similar to those of the Mini-Mental State Examination [38,39].

Usability testing is an important phase in the development of digital health interventions [34,35]. End users test a prototype in iterative cycles; they provide feedback about what works, what does not work, and where gaps might exist in the information and functionality [34,35]. These factors contribute to the frequency of use, understanding, and acceptability and enhance the likelihood that users will use the end product [35,36]. Testing the usability of the intervention also serves to assess the suitability of the platform interface and content [37].

Ethical Considerations

The phase-2B usability testing was approved by the Health Sciences Research Ethics Board (REB; 36415) at the University of Toronto on November 26, 2018, and the Health Sciences and Affiliated Teaching Hospitals REB (6026830) at Queen's University on June 20, 2019. An amendment was granted by the Health Sciences REB (36415) at the University of Toronto on March 26, 2020, to use secure video- or web conferencing (Zoom; Zoom Video Communications) for the usability testing sessions due to COVID-19 pandemic restrictions, which prohibited in-person participant contact. Informed consent and a demographic and clinical information form (Multimedia Appendix 1) were obtained from participants before each iterative cycle. Participants were identified using a study ID, and data were stored on a secure OneDrive folder (Microsoft Corp) accessible only to the principal investigator and research officer. All participants were compensated with a user-identified CAD $25 (US $17.52) gift card.

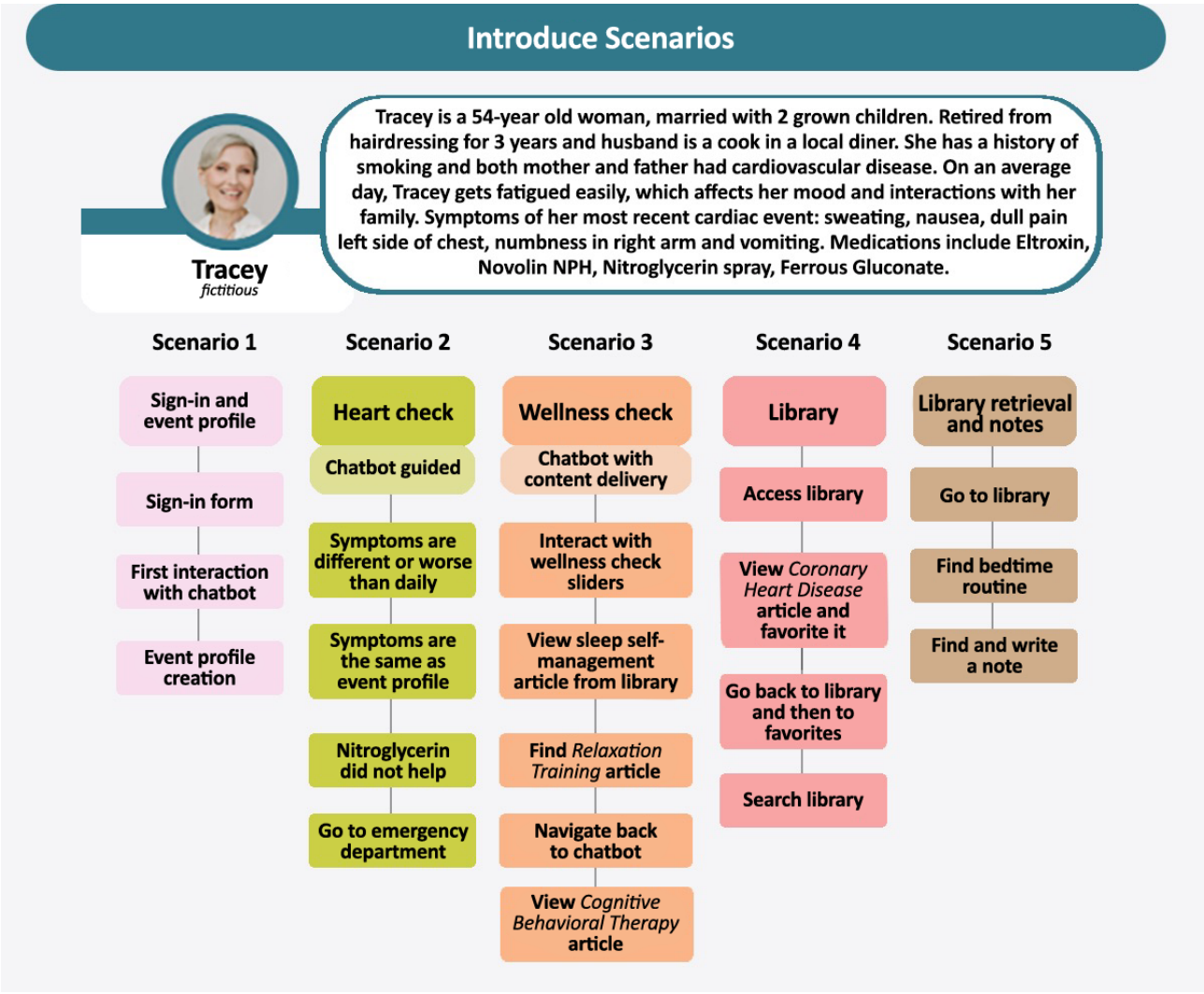
Study Design and Procedures

The usability of the *at heart* progressive web application focused on a think-aloud scenario-based approach to assess user performance (ie, ease of use, efficiency, and errors) and user satisfaction with *at heart's* content and functionality assessed qualitatively through the short scenario debrief questions and the interviews and quantitatively using the SUS [40]. This multi-method approach to usability testing is supported by a recent scoping review [41]. A sample of 10 eligible women was recruited to participate in the iterative usability testing cycles (n=5, 50% of the women participated in each of 2 testing cycles). This sample size was based on the experience of others [42-44], as well as recommendations that usability testing by 3 to 5 users can find approximately 85% of interface usability problems [45,46]. The first cycle tested user performance and satisfaction on a desktop computer, and the second cycle assessed user performance and satisfaction across various Android and iOS devices (ie, smartphones and tablets). Participants were provided with a brief explanation of the study and the *at heart* progressive web application before undergoing a 60- to 90-minute one-on-one observation period conducted and recorded through

the Zoom videoconferencing platform. Participants were introduced to a case of a woman aged 54 years with symptoms that were "possibly cardiac" [47] and asked to sign in, complete an event profile (scenario 1), and then progress through a set of standardized scenarios that incorporated each core feature of *at heart*: Heart Check, Wellness Check, and the library (scenarios 2-5; Figure 2).

The "think-aloud" approach was used to capture users' thought process and problem-solving as they progressed through *at heart's* core features in a systematic manner. After each scenario, the participants answered two short debrief questions: (1) "What did you like about the content and functionality of this specific scenario?" (2) "What did you dislike about the content and functionality of this specific scenario?"

Figure 2. Usability testing workflow.



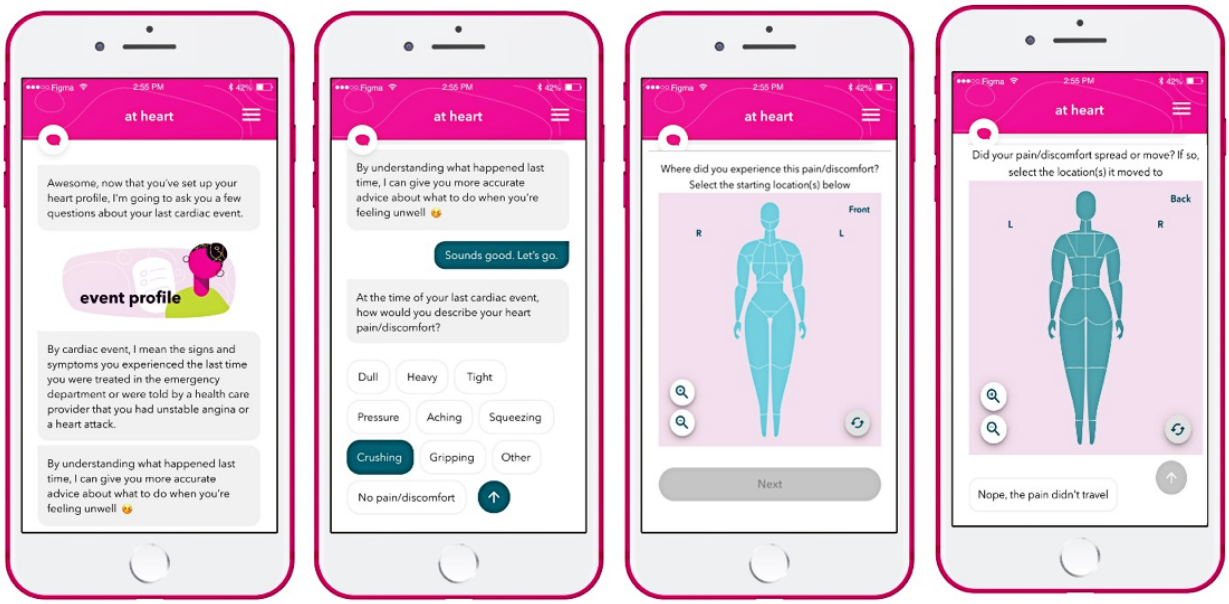
In total, 2 investigators also recorded user comments and documented any problems encountered during each scenario using the usability testing error and efficiency documentation form (Multimedia Appendix 2). Three error types were captured on this form: (1) navigation errors (ie, failure to locate a function or follow the recommended screen flow), (2) presentation errors (ie, selection errors due to labeling ambiguities), and (3) control use errors (ie, improper entry field errors). Four semistructured interview questions were posed following the think-aloud scenarios to assess the overall satisfaction with *at heart's* content and functionality: (1) "What was your overall impression of the *at heart* progressive web application?" (2) "What did you like or dislike about the *at heart* progressive web application?" 3) "Is there anything that could be improved or changed?" (4) "Is there anything missing from the *at heart* progressive web application?" Participants then completed the SUS on the web using REDCap (Research Electronic Data Capture; Vanderbilt University), a secure web application. The SUS is a widely used scale that quantifies the usability of digital health applications. It consists of 10 questions with Likert scales that range from *strongly agree* to *strongly disagree*, and it has been validated across a range of interfaces, including web pages and web applications [40,48,49]. Scores range from 0 to 100, with an accepted benchmark mean SUS score of 68 (SD 12.5) [40]. One 7-point adjective-anchored Likert scale was added to the bottom of the SUS so that participants could rate the overall user-friendliness of *at heart* as *worst imaginable*, *awful*, *poor*, *ok*, *good*, *excellent*, or *best imaginable*. This adjective rating scale helps inform the absolute usability of a product [50-52], such as the *at heart* progressive web application. Numerical equivalents of 1 (*worst imaginable*) to 7 (*best imaginable*) were assigned to the adjectives for scoring. After the first cycle,

revisions were made to the prototype, and the revised prototype was tested in a second usability testing cycle.

At Heart Progressive Web Application

At heart is a novel progressive web application that consists of 3 core features: Heart Check, Wellness Check, and a library. Users are guided through the core feature set by a rule-based chatbot that manages content and conversations. An Event Profile is created on the initial log-in to the progressive web application (Figure 3). The Event Profile contains individualized or personalized data that include the quality and location of cardiac pain or associated symptoms experienced at the time of the participant's last heart event (ie, treated in the emergency department). Female front and back full-body maps were specifically developed for *at heart* using the chest pain or associated symptom locations most commonly described in the literature [6,8,13,47,53].

Figure 3. Screenshots of the *at heart* feature Event Profile prototype.

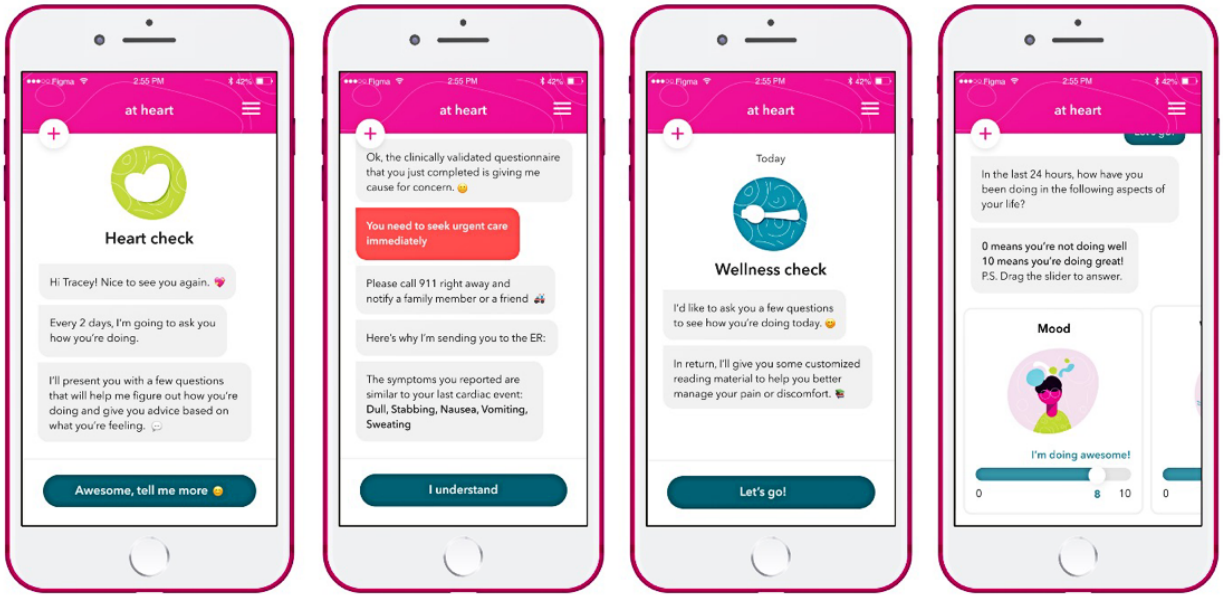


Chatbot Holly asks women the same series of cardiac pain and symptom assessment questions in the Heart Check, which is programmed into the *at heart* progressive web application to occur every 3 days (Figure 2). The Heart Check can occur more frequently simply by asking chatbot Holly for a Heart Check. Rule-based symptom triage algorithms compare each Heart Check to the individually stored Event Profile to make recommendations on an appropriate level of care. Level of care recommendations are based on the similarity and timing of current cardiac pain or discomfort to that stored Event Profile, noting similarities between the current and previous descriptions of the quality (ie, heavy, tight, or pressure), location, radiation, and associated symptoms. Recommendations are assigned to one of three categories: (1) *red* or high risk (ie, similar or high-risk symptoms that have occurred within 24 hours), (2) *yellow* or moderate risk (ie, similar or high-risk symptoms that have occurred beyond 24 hours but within 7 days), and (3) *green* or low risk (ie, no similar or high-risk symptoms). Women who are categorized as *red* are encouraged to notify a family member of their current symptoms and call 911 (ie, to seek appropriate and urgent assessment of their symptoms). Women who are categorized as *yellow* are encouraged to see their primary care provider within 48 hours, and women who are categorized as *green* are permitted access to the *at heart* progressive web application. High-risk symptoms are defined as ≥3 new typical symptom features (ie, dull, heavy, or tight chest pain) [53] or any new associated symptoms, including shortness of breath; palpitations; a racing heart rate; or feeling lightheaded, faint, or dizzy [47].

Women complete the Wellness Check, informed using features of the Brief Pain Inventory–Interference Subscale [54], to indicate the degree to which cardiac pain or associated symptoms interfere with 7 domains of life, which include general activities, paid and unpaid work, walking, mood, relations with others, sleep, and overall enjoyment of life (Figure 4). Wellness Checks are programmed to occur every 7 days at a minimum, and if the score in any domain is below a threshold, the chatbot uses rule-based algorithms to deliver educational content from the library.

The library contains scientific papers, each with a lay summary, covering educational content about CAD in women, with self-management advice to promote good health and well-being [29]. The library also contains videos and podcasts from women with lived experiences of CAD. *At heart* was evaluated by Health Canada and deemed not to be a medical device as defined in the Guidance Document: Software as a Medical Device (SaMD): Classification Examples [55]. *At heart* provides self-management support to women who have CAD and, through a chatbot, guides women to the most appropriate form of assessment (ie, primary care and emergent care) based on their medical symptoms.

Figure 4. Screenshots of the Heart Check and Wellness Check features of the *at heart* prototype.



Data Analysis

Quantitative descriptive data from the demographic and clinical information form and the SUS were analyzed using SPSS Statistics (version 29; IBM Corp). Debrief feedback from each of the 5 scenarios with data from the usability testing error and efficiency documentation form was reviewed after each of the 2 iterative usability testing cycles. Conventional content analysis as outlined by Sandelowski [56] was used to analyze qualitative data obtained from the interview questions. In total, 2 researchers independently assigned codes to label key actions, thoughts, and ideas. Related codes were grouped into categories, and any disagreements were discussed between the 2 researchers. A list of prototype issues was generated and prioritized based on risk to patients and ease of improvement [57]. Modifications were communicated to the website design and development team. Prototype changes made after the first cycle were evaluated in the second cycle, and prototype changes following the second cycle were completed before the phase-3 RCT.

Results

Participant Characteristics

A total of 10 women meeting the eligibility criteria were invited to participate in the usability testing from March 31, 2020, to April 17, 2020 (cycle 1), and from November 17, 2020, to November 30, 2020 (cycle 2). Women across the usability testing cycles had a mean age of 55.6 (SD 7.3) years, and most had a diploma (5/10, 50%), an undergraduate degree (2/10, 20%), or a graduate degree (2/10, 20%). Most of the women were employed full (3/10, 30%) or part time (2/10, 20%) and earned >CAD $70,000 (US $48,881.80) annually (6/10, 60%); 50% (5/10) lived with a disability, 50% (5/10) were the primary household earners, and 80% (8/10) were the primary persons responsible for housework in the home. The women reported cardiac pain or associated symptoms for 1 to 2 years (3/10, 30%), 2 to 5 years (3/10, 30%), or >5 years (3/10, 30%). The most common comorbid conditions included depression (4/10, 40%) and anxiety (4/10, 40%), and 40% (4/10) of the women had had a previous myocardial infarction. Participant demographics by usability testing cycle are reported in Table 1.

All women (10/10, 100%) had a computer at home and were comfortable (3/10, 30%) or very comfortable (7/10, 70%) using a computer, comfortable (1/10, 10%) or very comfortable (9/10, 90%) using the internet, and comfortable (3/10, 30%) or very comfortable (7/10, 70%) using smartphone or tablet apps. Participants' comfort levels and use of the computer, the internet, and smartphone or tablet apps are reported in Table 2. All women selected their own devices for usability testing. Women in the first cycle used only the Google Chrome browser (5/5, 100%), and women in the second cycle used an iPad mini version 9.3.5 (1/5, 20%), Samsung Galaxy S10e (1/5, 20%), iPhone 8 (2/5, 40%), and iPhone 11 (1/5, 20%).

Table 1. Participant characteristics.

Characteristic	Cycle 1 (n=5)	Cycle 2 (n=5)
Age (y), mean (SD)	58.4 (4.827)	52.8 (8.872)
Gender identity (woman), n (%)	5 (100)	5 (100)
Indigenous, n (%)	0 (0)	0 (0)
Identifying as a visible minority, n (%)	1 (20)	1 (20)
Living with a disability, n (%)	2 (40)	3 (60)
Gendered roles (housework hours), mean (SD)	6.6 (5.1)	10.2 (8.0)
Primary person responsible for housework, n (%)	4 (80)	4 (80)
Educational level, n (%)		
High school	0 (0)	1 (20)
Diploma	2 (40)	3 (60)
Bachelor's degree	2 (40)	0 (0)
Master's degree	1 (20)	1 (20)
Employment status, n (%)		
Full time	1 (20)	2 (40)
Part time	1 (20)	1 (20)
Unemployed	3 (60)	2 (40)
Income, n (%)		
<CAD $15,000 (US $10,474.70)	0 (0)	1 (20)
CAD $15,000-$29,900 (US $10,474.70-$20,879.50)	0 (0)	0 (0)
CAD $30,000-$49,900 (US $20,949.40-$34,845.80)	0 (0)	1 (20)
CAD $50,000-$69,900 (US $34,915.60-$48,812)	1 (20)	0 (0)
CAD $70,000-$99,900 (US $48,881.80-$69,761.40)	3 (60)	1 (20)
>CAD $100,000 (US $69,831.20)	1 (20)	1 (20)
Not reported	0 (0)	1 (20)
Primary earner, n (%)	1 (20)	4 (80)

Table 2. Use of the computer, the internet, and smartphone or tablet apps.

Characteristic	Cycle 1 (n=5)	Cycle 2 (n=5)
Computer use—home, n (%)	5 (100)	5 (100)
Computer use—work, n (%)	2 (40)	3 (60)
Computer hours per week, mean (SD)	6.4 (2.302)	8 (0.000)
Comfortable using computers, n (%)		
Comfortable	1 (20)	2 (40)
Very comfortable	4 (80)	3 (60)
Internet use, n (%)	5 (100)	5 (100)
Internet hours per week, mean (SD)	6.8 (2.168)	8 (0.000)
Comfortable using the internet, n (%)		
Comfortable	1 (20)	0 (0)
Very comfortable	4 (80)	5 (100)
Smartphone or tablet app use, n (%)	5 (100)	5 (100)
Smartphone or tablet app hours per week, mean (SD)	6.4 (2.302)	7 (1.414)
Comfortable using smartphone or tablet apps, n (%)		
Comfortable	1 (20)	2 (40)
Very comfortable	4 (80)	3 (60)

User Performance

User performance was assessed through ease of use, efficiency, and observation of testing errors (ie, navigation, presentation, and control use errors) through each scenario and cycle of usability testing. Testing errors included those related to navigation (difficulties moving through or locating content), presentation (selection errors due to labeling), or control use (improper entry field errors) across the 5 scenarios in 2 usability testing cycles (Figure 2). In scenario 1 of cycle 1 (sign-in, chatbot, and Event Profile), participants reported a low contrast between the text and the background and small font at sign-in. Participants also wanted clarification that the chatbot was not a real person. These were flagged as high priority, and subsequent revisions included larger font and chatbot and user text boxes identified using different colors, and a page was added to the *at heart* progressive web application to describe the chatbot, including its development and function. Specific text included the following: “let me tell you a little more about myself. I’m an educated chatbot, thoughtfully designed by a group of women’s heart health experts. I’ll be here to help you understand more about your heart pain/discomfort through regular check-ins, and I’ll send you tips and tricks for managing that pain. If you want to learn more about my makers, the same ones who named me Holly, go to the menu above...” A summary of suggestions and subsequent changes made to the progressive web application during usability testing cycle 1 is shown in Table 3.

The scenarios were repeated in the second cycle to assess user performance and user satisfaction across various Android and iOS devices (ie, smartphones and tablets). In scenario 1 of cycle 2 (sign-in, chatbot, and Event Profile), participants requested clarification on the terminology used in some of the questions at sign-in, so descriptors and more answers were added to these questions. Participants also had difficulty choosing specific areas on the body map, so the body maps were subsequently amended. In scenario 2 of cycle 2 (chatbot-guided workflow for the Heart Check), participants appeared confused with the order and speed of the questions delivered by the chatbot. The chat time stamp was subsequently increased for desktop, Android, and iOS devices. This allowed women ample time to read and respond appropriately to the Heart Check questions, especially those with longer messages and prompts. Finally, the Heart Check button was reported to be too small and gray, so the size and contrast were adjusted. The chatbot-guided workflow for the Wellness Check (scenario 3, cycle 2) required minor adjustments to ensure that the chatbot selected appropriate resources from the Wellness Library based on wellness scores of ≤4. One participant suggested that “anaphylaxis” be added to the Share My Data section and related articles be added to the library. As a result, lay summaries were generated, and scientific papers were added to the library for Kounis syndrome (n=3) and mast cell activation syndrome (n=2). Kounis syndrome is a complex multisystem arterial condition caused by mast cell activation and T-lymphocyte and macrophage interactions [58]. The resulting allergic, hypersensitivity, and anaphylactic reactions cause coronary symptoms (ie, cardiac pain), and this is referred to as Kounis syndrome [59]. A summary of suggestions and subsequent changes made to the progressive web application during usability testing cycle 2 is also shown in Table 3.

Table 3. Summary of suggestions and subsequent changes during usability testing.

Suggestions	Changes
Cycle 1 (n=5)	
Scenario 1—sign-in, chatbot, and Event Profile	
There was a low contrast between the text and the background and small font at sign-in.	Larger font and chatbot and user text boxes identified using different colors were implemented.
Clarification that the chatbot is not a real person was requested.	A page was added to the *at heart* progressive web application to describe the chatbot, including its development and function.
Participants selected only 1 descriptor on the body map and reported lack of clarity that more than one descriptor could be selected.	Directions for completing the body map were enhanced.
Scenario 2—Heart Check	
None	None
Scenario 3—Wellness Check	
None	None
Scenario 4—library	
Participants had difficulty locating the Favorites section of the library.	The Favorites section was moved to the top of the screen and identified with a larger icon that represented a heart.
Scenario 5—library retrieval and notes	
None	None
Other	
None	None
Cycle 2 (n=5)	
Scenario 1—sign-in, chatbot, and Event Profile	
Participants requested clarification on the terminology used in some of the questions at sign-in.	Descriptors and more answers were added to the questions for clarity.
Participants had difficulty choosing specific areas on the body map to describe their symptoms.	The body maps were subsequently amended to include axillae and body map descriptors.
Scenario 2—Heart Check	
Participants appeared confused with the order and speed of the questions delivered by the chatbot.	The chat speed was reduced for desktop, Android, and iOS devices.
The Heart Check button was reported to be too small and gray.	The size and contrast of the “Heart Check” button were adjusted.
Scenario 3—Wellness Check	
The chatbot says the following: “based on what you’ve told me, here are some topics...” However, the chatbot presented all topics even for wellness scores that were excellent (ie, the readings from the library did not appear tailored to the needs of each participant).	Minor adjustments were made to the rules to ensure that the chatbot selected appropriate resources from the wellness library based on wellness scores of ≤4.
Scenario 4—library	
None	None
Scenario 5—library retrieval and notes	
None	None
Other	
Add anaphylaxis to the Share My Data section.	This was added.
Add anaphylaxis articles to the library as they related to women and heart disease.	Lay summaries were generated and scientific papers were added to the library for Kounis syndrome (n=3) and mast cell activation syndrome (n=2).
Participants requested to have a more robust Share My Data section, with opportunity to record medications and other comorbid conditions.	The introduction and the content of the Share My Data section were enhanced.

User Satisfaction

Overview

User satisfaction was assessed using the SUS and the 4 semistructured interview questions. Overall, 10 participants across 2 cycles of testing reported the overall usability of the *at heart* progressive web application as highly acceptable (mean SUS score 81.75, SD 10.41). There was no difference in SUS scores between cycle 1 (mean SUS score 81.5, SD 8.59) and cycle 2 (mean SUS score 82.0, SD 13.04; t_8=−0.72, 95% CI −16.6 to 15.6). A total of 90% (9/10) of the participants rated the user-friendliness of *at heart* as good or excellent. All participants (10/10, 100%) thought that *at heart* was easy to use (SUS question 3; ie, most women could learn to use *at heart* very quickly):

> *I think everything was pretty straightforward. It was overall really, really good.* [Cycle 2; participant 3]

None of the participants indicated that they would need the support of a technical person to be able to use *at heart* (SUS question 4), and 80% (8/10) of the participants found the various functions in *at heart* well integrated (SUS question 5) with minimal inconsistencies (SUS question 6):

> *Well...I was a bit worried it might be confusing, but it's not confusing at all. It's very intuitive, it, you've built it so it's similar to a lot of other apps out there, right? So, clicking here, you know, submitting the button. So, it seems to be pretty normal as far, it looks, the same look and feel as other apps that are out there which are good. It seems to be easy. I'm not a tech person, so I wanted to make sure that—I was nervous that I wasn't going to be able to manage it because I, I'm not great with technology, but it was easy. So, yeah. I mean, I think it's a great tool.* [Cycle 2; participant 5]

A total of 8 themes emerged from the interview data (ie, engaging, comprehensive, understandable, credible, relevant, affirming, personalized, and innovative) and can be found in Multimedia Appendix 3.

Engaging

Participants commented on the layout, visual appeal, language, and name and logo of the progressive web application. They found the content easy to read and maneuver. Important content was highlighted, and participants reported this to be helpful. Article summaries were presented in larger font and used lay language. At the outset of this study, the progressive web application was named HEARTPA♀N [29]. Participants recommended a name change to one that was less focused on pain as many women describe their cardiac pain as discomfort and have other associated symptoms (eg, dyspnea and fatigue). They felt that the name *at heart* was more personally welcoming and less focused on pain. The heart logo was simple yet distinctive and recognizable.

Comprehensive

Participants indicated that the progressive web application provided necessary information for helping them make decisions about their cardiac symptoms. They liked that it was specific to women and that they could search through the library to improve their knowledge about wellness and women's heart health. Participants also commented that the chatbot assisted them in making decisions:

> *...the bot kind of identifies for me what the big blind spots are that I am perhaps not noticing.*

Understandable

Participants indicated that *at heart* was intuitive, the content made sense, and it flowed well. The tone and the language were suitable and clear. Participants commented on the appropriateness of the associated symptom descriptors included in the Event Profile and Heart Check.

Credible

Participants liked that *at heart* was developed as a living progressive web application. They particularly liked the breadth and depth of scientific articles and lay summaries included in the *at heart* library. Participants valued having an evidence-informed digital health self-management program specifically designed for women with CAD living with cardiac pain and associated symptoms.

Relevant

Participants commented on the relevance and accessibility of content in the *at heart* library. The chatbot asked relevant questions and was able to use rule-based algorithms to provide advice to women about timely assessment in the emergency department. They particularly noted the relevance of having their Event Profile and Heart Check information stored so that they could refer to it when visiting their health care providers. The content was applicable to their everyday lives.

Affirming

Participants looked for affirmation from *at heart's* chatbot when they were having cardiac symptoms. They indicated that they would have otherwise consulted with family members or Facebook group members for guidance. Affirmation of symptoms was important to women, and they appreciated the high-alert messages to seek urgent care immediately.

Personalized

Participants liked the chatbot. They found it to be personal, immediately interactive, accessible, calming, and friendly. They commented that the chatbot may help make them feel less alone if they went to the emergency room because it checked in with them to see how they were doing. The chatbot provided comfort at times when participants felt scared. Participants also valued the videos and podcasts as they also helped them feel that they were not alone in their experience with heart disease.

Innovative

Participants felt that *at heart* provided an opportunity for further research focused on women's heart health. They were willing to share anonymized data on socioeconomics, risk factors, gynecological or obstetrics history, lifestyle, medications, and other conditions (eg, arthritis or depression) to improve outcomes for women at risk of or with heart disease. Participants also valued the ability to journal and take notes in the

progressive web application. They valued the opportunities that *at heart* could provide to women who live in more isolated areas or those who live alone. The *at heart* progressive web application was described as all-encompassing, covering necessary content to improve knowledge and decision-making, with a chatbot to ask questions and deliver information in a comforting manner.

Discussion

Principal Findings

There is growing support for the importance of usability testing for mobile health and eHealth innovations [60-62]. Usability testing incorporates an iterative process of testing and refining to meet end-user needs [63]. The objectives of this study were to assess the user performance (ie, ease of use, efficiency, and errors) and user satisfaction (SUS) of a progressive web application for women with cardiac pain. User performance was assessed through ease of use, efficiency, and observation of testing errors (ie, navigation, presentation, and control use errors) through 5 brief scenarios and 2 cycles of usability testing.

The overall usability of the *at heart* progressive web application was rated highly by 10 participants who completed the usability testing cycles; the progressive web application was easy to use and efficient. However, 2 high-priority testing errors were identified during the usability testing cycles. The first involved small font at sign-in and low contrast between the text and background, and these were corrected by enlarging the font and making alterations to the text and background colors. The second was to clarify that the chatbot was not a real person. A significant amount of time was committed to addressing potential language challenges during the development of the chatbot, and there were no technical, design, or language challenges identified during the usability testing cycles. In fact, the design and language or conversation of the chatbot appeared to mimic human-to-human interaction so closely that participants requested confirmation that the chatbot was actually nonhuman.

The *at heart* chatbot is a simple rule-based conversational agent designed to mimic human-to-human interaction at sign-in and creation of an Event Profile and during completion of the Heart Check and Wellness Check. Our previous systematic review and meta-analysis indicated that self-management programs were more effective in reducing cardiac pain and associated symptoms (eg, dyspnea and fatigue) in women if they included collaboration and support from health care providers [33]. Chatbots are conversational agents able to promote health and provide education and support [64]. Their use in health care is still in a developmental stage; they could improve access to care and health care provider–patient communication, but more evidence is needed. Technical, design (ie, lack of empathy), and language challenges [65] can impede the integration of chatbot technologies into health care [64]. Adopting user-centered and theory-based designs, optimizing user experiences, and addressing patient concerns can improve chatbot uptake and use [64]. Our chatbot relies on scripted computational algorithms, with specific rules for its text-based conversations [66]. More advanced human-machine conversation is now based on natural language processing and large language models that use artificial intelligence methods to learn, understand, and produce structured language content [66,67]. Health care delivery is rapidly changing and is driven by social, scientific, and technological change; our future chatbot may need to emulate person-to-person conversation through dialogue and body movements, with appropriate expressions of empathy and compassion [68]. We are really at a cusp in health care; growth in chatbot use will be driven by a desire for health and wellness and 24-hour access to care, with a growing number of platforms from which to build an intelligent and emotive chatbot in the future [69].

At heart is the first of its kind; no smartphone- or web-based self-management program has been co-designed and systematically developed with women who have lived experience (phase 2A) and then tested with women who have cardiac pain (phase 2B). We used the individual and family self-management theory [70,71], mobile device functionality, and the pervasive information architecture of mobile health interventions [72] and followed the MRC's guidance for developing complex interventions [30-32]. *At heart* was designed as a progressive web application aimed at delivering nativelike user experiences regardless of the browser or the mobile device [73]. This study identified that the comprehensiveness and credibility of the information was important to women. The information helped them understand more about CAD and wellness and provided them with guidance in decision-making. At inception, women clearly articulated the need for a web application that was accessible across Android and iOS operating systems, including computers, smartphones, and tablets. In phase 1, women identified the need to access the library anytime and anywhere (eg, while waiting for dental appointments). As women were involved at the outset in co-design, no other significant user performance or satisfaction–related issues were identified. Female front and back full-body maps were specifically developed for *at heart* using the chest pain or associated symptom locations most commonly described in the literature [6,8,13,47,53]. These required only minor refinements during usability testing (ie, more precise identification of the axillae). Importantly, participants also viewed *at heart* as a way to contribute to the future of women's heart health. Women feel "stopped at the gate," and they want to take charge and advocate for better awareness, education, diagnosis, and management [74].

Limitations and Strengths

First, usability testing was conducted on a predominantly affluent sample of participants, which limits the generalizability of our results. A total of 60% (6/10) of the participants who took part in usability testing earned >CAD $70,000 (US $48,881.80) annually. They were also well educated, and most (8/10, 80%) were White. Canada is known for its diversity in race and ethnicities— over 450 ethnic or cultural origins were reported in the 2021 census [75]. A total of 1 out of 5 people in Canada is born elsewhere; the 3 largest visible ethnic groups are South Asian, Chinese, and Black individuals, representing 60% of the Canadian racial and ethnic populations [76]. Although this was a limitation in the sample for the usability testing, this is not a limitation specific to the *at heart* progressive web application. In fact, the *at heart* library contains heart

disease and heart wellness scientific papers and lay summaries related to South Asian and Indigenous people. These include risk and traditional practices such as drumming and healing and talking circles. However, it will be necessary to include a more diverse sample of women that encompasses those in less populous areas where age-standardized cardiovascular-related deaths in women are highest [77]. It will also be important to target recruitment of women from South Asian and Afro-Caribbean races and ethnicities, those with disabilities, those of a lower socioeconomic status, and those across age and gender identities who also experience greater risk of cardiovascular disease [77-79].

In addition, heterogeneous use of platforms and devices may also be a limitation to the usability testing. Progressive web applications are designed to work across platforms, eliminating the cost, fragmentation, and need to develop the same application several times for multiple platforms. They are designed to behave like mobile apps, working seamlessly across Android and iOS devices. However, past research indicates that progressive web applications may not be as smooth as mobile apps on Android devices [80]. Our sample size was limited, and only 10% (1/10) of the participants used an Android device for testing. Further research is needed to evaluate the usability of the *at heart* progressive web application across various platforms and devices. The phase-3 pilot RCT will determine the feasibility of implementing the *at heart* intervention, including the *feasibility* of randomization, recruitment, and retention; *acceptability* and barriers to implementing the intervention; and the extent of engagement with the intervention. The phase-4 effectiveness-implementation hybrid trial (type I) will investigate the effectiveness and implementation of *at heart* among women living in Canada using the practical, robust implementation and sustainability model [81]. The MRC [30-32] discusses 4 phases of complex intervention research (eg, development, feasibility, evaluation, and implementation) and the importance of refinements with each phase transition depending on the context [70,71] in which the intervention is evaluated and implemented. The practical, robust implementation and sustainability model [81] aligns contextual needs with intervention design using an inequity, systems thinking, and cocreation engagement approach. Context will be addressed by considering the individual perspectives on the intervention, external environment, characteristics of those receiving the intervention, and supports and resources needed to deliver the intervention [81]. Recruiting representative populations can be improved through strengthened community partnerships in governance and decision-making [82-84], and this will be a focus in subsequent phases of this research.

Second, remote usability testing methods were used during the COVID-19 pandemic. Testing a technology's usability remotely has inherent advantages as well as limitations. While participants can prefer remote testing to in-person testing, it can take longer for participants to complete tasks during remote testing, more errors can be made compared to in-person testing sessions, and context and indirect cues can be missed during remote testing sessions [85]. However, the approach taken to usability testing in this study was iterative, comprehensive, and aligned with methods used by others recently reported in a scoping review [41]. Of 133 articles included in this scoping review, 105 used questionnaires, 45 used a think-aloud approach, and 37 used interviews in their usability testing methods [41]. Moreover, many studies used a combination of 2 (n=46) or 3 (n=30) methods of testing, and the SUS was the most reported and validated questionnaire used to assess satisfaction and usability [41]. Finally, the *at heart* chatbot is rule based and designed for specific functions within the progressive web application. It is not led by artificial intelligence technologies (ie, natural language processing or large language models) and, therefore, may have inherent limitations that were not identified during the 2 cycles of usability testing (ie, personality, flexibility, dialogue structure, and conversation complexity or flow) [65].

Conclusions

This study does provide initial support for the *at heart* progressive web application for women with cardiac pain. Participants rated the performance of and satisfaction with this progressive web application as high. Ongoing evaluations in phases 3 and 4 should aim to examine the feasibility and acceptability and the extent of engagement with the *at heart* core feature set: Heart Check, Wellness Check, and the library. In addition to assessing effectiveness in an effectiveness-implementation hybrid trial (type I), describing and better understanding the context for implementation (eg, race and ethnicity and geography) will be necessary.

Acknowledgments

The authors thank the women who were participants in this study. They also thank the women with lived experience who collaborated in the co-design of *at heart* and the Heart and Stroke Foundation of Canada for their partnership. Phases 2 and 3 of the *at heart* program of research are supported by the Canadian Institutes of Health Research (389044) Project Grant: Fall 2017 competition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant demographic and clinical information form.
[PDF File (Adobe PDF File), 94 KB - humanfactors_v12i1e57583_app1.pdf]

Multimedia Appendix 2
Usability testing error and efficiency documentation form.
[PDF File (Adobe PDF File), 72 KB - humanfactors_v12i1e57583_app2.pdf]

Multimedia Appendix 3
Summary of participant comments by theme.
[PDF File (Adobe PDF File), 103 KB - humanfactors_v12i1e57583_app3.pdf]

References

1. GBD 2017 Causes of Death Collaborators. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788 [FREE Full text] [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. J Am Coll Cardiol 2020 Dec 22;76(25):2982-3021 [FREE Full text] [doi: 10.1016/j.jacc.2020.11.010] [Medline: 33309175]
3. Vaduganathan M, Mensah GA, Turco JV, Fuster V, Roth GA. The global burden of cardiovascular diseases and risk: a compass for future health. J Am Coll Cardiol 2022 Dec 20;80(25):2361-2371 [FREE Full text] [doi: 10.1016/j.jacc.2022.11.005] [Medline: 36368511]
4. Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. J Am Coll Cardiol 2019 Nov 19;74(20):2529-2532 [FREE Full text] [doi: 10.1016/j.jacc.2019.10.009] [Medline: 31727292]
5. Pepine CJ, Ferdinand KC, Shaw LJ, Light-McGroary KA, Shah RU, Gulati M, et al. Emergence of nonobstructive coronary artery disease: a woman's problem and need for change in definition on angiography. J Am Coll Cardiol 2015 Oct 27;66(17):1918-1933 [FREE Full text] [doi: 10.1016/j.jacc.2015.08.876] [Medline: 26493665]
6. McSweeney J, Cleves MA, Fischer EP, Moser DK, Wei J, Pettey C, et al. Predicting coronary heart disease events in women: a longitudinal cohort study. J Cardiovasc Nurs 2014;29(6):482-492 [FREE Full text] [doi: 10.1097/JCN.0b013e3182a409cc] [Medline: 24231895]
7. Parry M, Bjørnnes AK, Clarke H, Cooper L, Gordon A, Harvey P, et al. Self-management of cardiac pain in women: an evidence map. BMJ Open 2017 Nov 25;7(11):e018549 [FREE Full text] [doi: 10.1136/bmjopen-2017-018549] [Medline: 29175891]
8. Canto JG, Canto EA, Goldberg RJ. Time to standardize and broaden the criteria of acute coronary syndrome symptom presentations in women. Can J Cardiol 2014 Jul;30(7):721-728. [doi: 10.1016/j.cjca.2013.10.015] [Medline: 24530216]
9. Kirchberger I, Heier M, Wende R, von Scheidt W, Meisinger C. The patient's interpretation of myocardial infarction symptoms and its role in the decision process to seek treatment: the MONICA/KORA Myocardial Infarction Registry. Clin Res Cardiol 2012 Nov;101(11):909-916. [doi: 10.1007/s00392-012-0475-8] [Medline: 22618810]
10. Kok MM, van der Heijden LC, Sen H, Danse PW, Löwik MM, Anthonio RL, et al. Sex difference in chest pain after implantation of newer generation coronary drug-eluting stents: a patient-level pooled analysis from the TWENTE and DUTCH PEERS trials. JACC Cardiovasc Interv 2016 Mar 28;9(6):553-561 [FREE Full text] [doi: 10.1016/j.jcin.2015.10.043] [Medline: 26947384]
11. Choinière M, Watt-Watson J, Victor JC, Baskett RJ, Bussières JS, Carrier M, et al. Prevalence of and risk factors for persistent postoperative nonanginal pain after cardiac surgery: a 2-year prospective multicentre study. CMAJ 2014 Apr 15;186(7):E213-E223 [FREE Full text] [doi: 10.1503/cmaj.131012] [Medline: 24566643]
12. Katz J, Weinrib A, Fashler SR, Katznelzon R, Shah BR, Ladak SS, et al. The Toronto General Hospital Transitional Pain Service: development and implementation of a multidisciplinary program to prevent chronic postsurgical pain. J Pain Res 2015;8:695-702 [FREE Full text] [doi: 10.2147/JPR.S91924] [Medline: 26508886]
13. Sella YO, Manistamara H, Apriliawan S, Lukitasari M, Rohman MS. Characteristic differences of chest pain in male and female patients with acute coronary syndrome: a pilot study. J Public Health Res 2021 Apr 14;10(2):2242 [FREE Full text] [doi: 10.4081/jphr.2021.2242] [Medline: 33855424]
14. Altintas E, Yigit F, Taskintuna N. The impact of psychiatric disorders with cardiac syndrome X on quality of life: 3 months prospective study. Int J Clin Exp Med 2014;7(10):3520-3527 [FREE Full text] [Medline: 25419392]
15. Conradie A, Atherton J, Chowdhury E, Duong M, Schwarz N, Worthley S, et al. Health-Related Quality of Life (HRQoL) and the effect on outcome in patients presenting with coronary artery disease and treated with Percutaneous Coronary Intervention (PCI): differences noted by sex and age. J Clin Med 2022 Sep 05;11(17):5231 [FREE Full text] [doi: 10.3390/jcm11175231] [Medline: 36079161]
16. Sabbag A, Matetzky S, Gottlieb S, Fefer P, Kohanov O, Atar S, et al. Recent temporal trends in the presentation, management, and outcome of women hospitalized with acute coronary syndromes. Am J Med 2015 Apr;128(4):380-388. [doi: 10.1016/j.amjmed.2014.10.041] [Medline: 25460528]

17. Sjöström-Strand A, Fridlund B. Women's descriptions of symptoms and delay reasons in seeking medical care at the time of a first myocardial infarction: a qualitative study. Int J Nurs Stud 2008 Jul;45(7):1003-1010. [doi: 10.1016/j.ijnurstu.2007.07.004] [Medline: 17803997]
18. Hadid LA, Al Barmawi M, Al Hmaimat NA, Shoqirat N. Factors associated with prehospital delay among men and women newly experiencing acute coronary syndrome: a qualitative inquiry. Cardiol Res Pract 2020;2020:3916361 [FREE Full text] [doi: 10.1155/2020/3916361] [Medline: 32454998]
19. Lorig KR, Holman H. Self-management education: history, definition, outcomes, and mechanisms. Ann Behav Med 2003 Aug;26(1):1-7. [doi: 10.1207/S15324796ABM2601_01] [Medline: 12867348]
20. Osborne RH, Batterham R, Livingston J. The evaluation of chronic disease self-management support across settings: the international experience of the health education impact questionnaire quality monitoring system. Nurs Clin North Am 2011 Sep;46(3):255-70, v. [doi: 10.1016/j.cnur.2011.05.010] [Medline: 21791261]
21. Barlow J, Wright C, Sheasby J, Turner A, Hainsworth J. Self-management approaches for people with chronic conditions: a review. Patient Educ Couns 2002;48(2):177-187. [doi: 10.1016/s0738-3991(02)00032-0] [Medline: 12401421]
22. Knight-Agarwal C, Davis DL, Williams L, Davey R, Cox R, Clarke A. Development and pilot testing of the Eating4two mobile phone app to monitor gestational weight gain. JMIR Mhealth Uhealth 2015 Jun 05;3(2):e44 [FREE Full text] [doi: 10.2196/mhealth.4071] [Medline: 26048313]
23. Willcox JC, Campbell KJ, McCarthy EA, Wilkinson SA, Lappas M, Ball K, et al. Testing the feasibility of a mobile technology intervention promoting healthy gestational weight gain in pregnant women (txt4two) - study protocol for a randomised controlled trial. Trials 2015 May 07;16:209 [FREE Full text] [doi: 10.1186/s13063-015-0730-1] [Medline: 25947578]
24. Waring ME, Moore Simas TA, Xiao RS, Lombardini LM, Allison JJ, Rosal MC, et al. Pregnant women's interest in a website or mobile application for healthy gestational weight gain. Sex Reprod Healthc 2014 Dec;5(4):182-184 [FREE Full text] [doi: 10.1016/j.srhc.2014.05.002] [Medline: 25433828]
25. Fukuoka Y, Komatsu J, Suarez L, Vittinghoff E, Haskell W, Noorishad T, et al. The mPED randomized controlled clinical trial: applying mobile persuasive technologies to increase physical activity in sedentary women protocol. BMC Public Health 2011 Dec 14;11:933. [doi: 10.1186/1471-2458-11-933] [Medline: 22168267]
26. Llanos AA, Krok JL, Peng J, Pennell ML, Vitolins MZ, Degraffinreid CR, et al. Effects of a walking intervention using mobile technology and interactive voice response on serum adipokines among postmenopausal women at increased breast cancer risk. Horm Cancer 2014 Apr;5(2):98-103 [FREE Full text] [doi: 10.1007/s12672-013-0168-4] [Medline: 24435584]
27. Wen KY, Miller SM, Kilby L, Fleisher L, Belton TD, Roy G, et al. Preventing postpartum smoking relapse among inner city women: development of a theory-based and evidence-guided text messaging intervention. JMIR Res Protoc 2014 Apr 03;3(2):e20 [FREE Full text] [doi: 10.2196/resprot.3059] [Medline: 24698804]
28. Licskai C, Sands TW, Ferrone M. Development and pilot testing of a mobile health solution for asthma self-management: asthma action plan smartphone application pilot study. Can Respir J 2013;20(4):301-306 [FREE Full text] [doi: 10.1155/2013/906710] [Medline: 23936890]
29. Parry M, Dhukai A, Clarke H, Bjørnnes AK, Cafazzo JA, Cooper L, et al. Development and usability testing of HEARTPA ♀ N: protocol for a mixed methods strategy to develop an integrated smartphone and web-based intervention for women with cardiac pain. BMJ Open 2020 Mar 09;10(3):e033092 [FREE Full text] [doi: 10.1136/bmjopen-2019-033092] [Medline: 32156763]
30. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
31. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. Health Technol Assess 2021 Sep;25(57):1-132 [FREE Full text] [doi: 10.3310/hta25570] [Medline: 34590577]
32. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ 2000 Sep 16;321(7262):694-696 [FREE Full text] [doi: 10.1136/bmj.321.7262.694] [Medline: 10987780]
33. Parry M, Bjørnnes AK, Victor JC, Ayala AP, Lenton E, Clarke H, et al. Self-management interventions for women with cardiac pain: a systematic review and meta-analysis. Can J Cardiol 2018 Apr;34(4):458-467. [doi: 10.1016/j.cjca.2017.12.011] [Medline: 29477931]
34. Wichansky AM. Usability testing in 2000 and beyond. Ergonomics 2000 Jul;43(7):998-1006. [doi: 10.1080/001401300409170] [Medline: 10929833]
35. Currie LM. Evaluation frameworks for nursing informatics. Int J Med Inform 2005 Dec;74(11-12):908-916. [doi: 10.1016/j.ijmedinf.2005.07.007] [Medline: 16099711]
36. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pract 2003 Oct;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
37. Gustafson DH, Wyatt JC. Evaluation of ehealth systems and services. BMJ 2004 May 15;328(7449):1150. [doi: 10.1136/bmj.328.7449.1150] [Medline: 15142895]

38. Callahan CM, Unverzagt FW, Hui SL, Perkins AJ, Hendrie HC. Six-item screener to identify cognitive impairment among potential subjects for clinical research. Med Care 2002 Sep;40(9):771-781. [doi: 10.1097/00005650-200209000-00007] [Medline: 12218768]
39. Carpenter CR, DesPain B, Keeling TN, Shah M, Rothenberger M. The Six-Item Screener and AD8 for the detection of cognitive impairment in geriatric emergency department patients. Ann Emerg Med 2011 Jun;57(6):653-661 [FREE Full text] [doi: 10.1016/j.annemergmed.2010.06.560] [Medline: 20855129]
40. Hyzy M, Bond R, Mulvenna M, Bai L, Dix A, Leigh S, et al. System usability scale benchmarking for digital health apps: meta-analysis. JMIR Mhealth Uhealth 2022 Aug 18;10(8):e37290 [FREE Full text] [doi: 10.2196/37290] [Medline: 35980732]
41. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
42. Fleming JB, Hill YN, Burns MN. Usability of a culturally informed mHealth intervention for symptoms of anxiety and depression: feedback from young sexual minority men. JMIR Hum Factors 2017 Aug 25;4(3):e22. [doi: 10.2196/humanfactors.7392] [Medline: 28842389]
43. Srikesavan C, Williamson E, Cranston T, Hunter J, Adams J, Lamb SE. An online hand exercise intervention for adults with rheumatoid arthritis (mySARAH): design, development, and usability testing. J Med Internet Res 2018 Jun 27;20(6):e10457 [FREE Full text] [doi: 10.2196/10457] [Medline: 29950288]
44. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de Vries H, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013 Jan 16;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
45. Macefield R. How to specify the participant group size for usability studies: a practitioner's guide. J Usab Stud 2009:34-45 [FREE Full text]
46. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. 1993 May 01 Presented at: CHI '93; April 24-29, 1993; Amsterdam, The Netherlands URL: https://dl.acm.org/doi/abs/10.1145/169059.169166 [doi: 10.1145/169059.169166]
47. Gulati M, Levy PD, Mukherjee D, Amsterdam E, Bhatt DL, Birtcher KK, et al. 2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR guideline for the evaluation and diagnosis of chest pain: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. Circulation 2021 Nov 30;144(22):e368-e454 [FREE Full text] [doi: 10.1161/CIR.0000000000001029] [Medline: 34709879]
48. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996.
49. Peres SC, Pham T, Phillips R. Validation of the system usability scale (SUS): SUS in the wild. Proc Hum Factors Ergon Soc Annu Meet 2013 Sep 30;57(1):192-196. [doi: 10.1177/1541931213571043]
50. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
51. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
52. Gräve E, Buchner A. Is less sometimes more? An experimental comparison of four measures of perceived usability. Hum Factors 2025 Jan;67(1):32-48 [FREE Full text] [doi: 10.1177/00187208241237862] [Medline: 38482806]
53. Ferry AV, Anand A, Strachan FE, Mooney L, Stewart SD, Marshall L, et al. Presenting symptoms in men and women diagnosed with myocardial infarction using sex‐specific criteria. JAHA 2019 Sep 03;8(17). [doi: 10.1161/jaha.119.012307]
54. Gjeilo KH, Stenseth R, Wahba A, Lydersen S, Klepstad P. Validation of the brief pain inventory in patients six months after cardiac surgery. J Pain Symptom Manage 2007 Dec;34(6):648-656 [FREE Full text] [doi: 10.1016/j.jpainsymman.2007.01.010] [Medline: 17629665]
55. Health Canada. Guidance document: Software as a Medical Device (SaMD): classification examples. Government of Canada. Ottawa, ON: Health Canada; 2019 Dec 18. URL: https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/medical-devices/application-information/guidance-documents/software-medical-device-guidance/examples/examples.pdf [accessed 2025-03-26]
56. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [doi: 10.1002/1098-240x(200008)23:4<334::aid-nur9>3.0.co;2-g] [Medline: 10940958]
57. Rogers ML, Patterson E, Chapman R, Render M. Usability testing and the relation of clinical information systems to patient safety. In: Henriksen K, Battles JB, Marks ES, Lewin DI, editors. Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology). Rockville, MD: Agency for Healthcare Research and Quality (US); 2005.
58. Kounis NG. Coronary hypersensitivity disorder: the Kounis syndrome. Clin Ther 2013 May;35(5):563-571. [doi: 10.1016/j.clinthera.2013.02.022] [Medline: 23490289]
59. Kounis NG, Koniari I, Velissaris D, Tzanis G, Hahalis G. Kounis syndrome—not a single-organ arterial disorder but a multisystem and multidisciplinary disease. Balkan Med J 2019 Jul 11;36(4):212-221 [FREE Full text] [doi: 10.4274/balkanmedj.galenos.2019.2019.5.62] [Medline: 31198019]

60. Hales S, Turner-McGrievy G, Fahim A, Freix A, Wilcox S, Davis RE, et al. A mixed-methods approach to the development, refinement, and pilot testing of social networks for improving healthy behaviors. JMIR Hum Factors 2016 Feb 12;3(1):e8 [FREE Full text] [doi: 10.2196/humanfactors.4512] [Medline: 27026535]
61. Kaufman DR, Patel VL, Hilliman C, Morin PC, Pevzner J, Weinstock RS, et al. Usability in the real world: assessing medical information technologies in patients' homes. J Biomed Inform 2003 Feb;36(1-2):45-60 [FREE Full text] [doi: 10.1016/s1532-0464(03)00056-x] [Medline: 14552846]
62. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Comput Biol Med 2002 May;32(3):141-149. [doi: 10.1016/s0010-4825(02)00011-2] [Medline: 11922931]
63. Snodgrass A, Coyne R. Models, metaphors and the hermeneutics of designing. Design Issues 1992;9(1):56-74. [doi: 10.2307/1511599]
64. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019 Aug 21;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
65. Fadhil A, Schiavo G. Designing for health chatbots. arXiv Preprint posted online on February 24, 2019 [FREE Full text]
66. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015 Jul 17;349(6245):261-266. [doi: 10.1126/science.aaa8685] [Medline: 26185244]
67. Lopez-Jimenez F, Attia Z, Arruda-Olson AM, Carter R, Chareonthaitawee P, Jouni H, et al. Artificial intelligence in cardiology: present and future. Mayo Clin Proc 2020 May;95(5):1015-1039. [doi: 10.1016/j.mayocp.2020.01.038] [Medline: 32370835]
68. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. J Med Syst 2019 Jun 25;43(8):246. [doi: 10.1007/s10916-019-1387-1] [Medline: 31240494]
69. Chakraborty C, Pal S, Bhattacharya M, Dash S, Lee SS. Overview of Chatbots with special emphasis on artificial intelligence-enabled ChatGPT in medical science. Front Artif Intell 2023;6:1237704 [FREE Full text] [doi: 10.3389/frai.2023.1237704] [Medline: 38028668]
70. Fawcett J, Watson J, Neuman B, Walker PH, Fitzpatrick JJ. On nursing theories and evidence. J Nurs Scholarsh 2001;33(2):115-119. [doi: 10.1111/j.1547-5069.2001.00115.x] [Medline: 11419305]
71. Ryan P, Sawin KJ. The individual and family self-management theory: background and perspectives on context, process, and outcomes. Nurs Outlook 2009 Jul;57(4):217-25.e6 [FREE Full text] [doi: 10.1016/j.outlook.2008.10.004] [Medline: 19631064]
72. Danaher BG, Brendryen H, Seeley JR, Tyler MS, Woolley T. From black box to toolbox: outlining device functionality, engagement activities, and the pervasive information architecture of mHealth interventions. Internet Interv 2015 Mar 01;2(1):91-101 [FREE Full text] [doi: 10.1016/j.invent.2015.01.002] [Medline: 25750862]
73. Tandel SS, Jamadar A. Impact of progressive web apps on web app development. Int J Innov Res Sci Eng Technol 2018;7(9):9439-9444. [doi: 10.15680/IJIRSET.2018.0709021]
74. Colella TJ, Hardy M, Hart D, Price JA, Sarfi H, Mullen KA, et al. The Canadian Women's Heart Health Alliance Atlas on the epidemiology, diagnosis, and management of cardiovascular disease in women-chapter 3: patient perspectives. CJC Open 2021 Mar;3(3):229-235 [FREE Full text] [doi: 10.1016/j.cjco.2020.11.020] [Medline: 33778439]
75. The Canadian census: a rich portrait of the country's religious and ethnocultural diversity. Statistics Canada. Ottawa, ON: Statistics Canada; 2022 Oct 26. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/221026/dq221026b-eng.htm [accessed 2024-01-23]
76. Immigration and ethnocultural diversity in Canada. Statistics Canada. Ottawa, ON: Statistics Canada; 2011. URL: https://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.cfm [accessed 2025-03-26]
77. Jaffer S, Foulds HJ, Parry M, Gonsalves CA, Pacheco C, Clavel MA, et al. The Canadian Women's Heart Health Alliance Atlas on the epidemiology, diagnosis, and management of cardiovascular disease in women-chapter 2: scope of the problem. CJC Open 2021 Jan;3(1):1-11 [FREE Full text] [doi: 10.1016/j.cjco.2020.10.009] [Medline: 33458627]
78. Norris CM, Yip CY, Nerenberg KA, Clavel M, Pacheco C, Foulds HJ, et al. State of the science in women's cardiovascular disease: a Canadian perspective on the influence of sex and gender. J Am Heart Assoc 2020 Feb 18;9(4):e015634. [doi: 10.1161/JAHA.119.015634] [Medline: 32063119]
79. Parry M, Van Spall HG, Mullen KA, Mulvagh SL, Pacheco C, Colella TJ, et al. The Canadian Women's Heart Health Alliance Atlas on the epidemiology, diagnosis, and management of cardiovascular disease in women - chapter 6: sex- and gender-specific diagnosis and treatment. CJC Open 2022 Jul;4(7):589-608 [FREE Full text] [doi: 10.1016/j.cjco.2022.04.002] [Medline: 35865023]
80. Fournier C. Comparison of smoothness in progressive web apps and mobile applications on android. KTH Royal Institute of Technology. 2020. URL: https://www.diva-portal.org/smash/get/diva2:1474729/FULLTEXT01.pdf [accessed 2025-03-25]
81. Jolles MP, Fort MP, Glasgow RE. Aligning the planning, development, and implementation of complex interventions to local contexts with an equity focus: application of the PRISM/RE-AIM Framework. Int J Equity Health 2024 Feb 27;23(1):41 [FREE Full text] [doi: 10.1186/s12939-024-02130-6] [Medline: 38408990]

82. Herington J, Connelly K, Illes J. Ethical imperatives for working with diverse populations in digital research. J Med Internet Res 2023 Sep 18;25:e47884 [FREE Full text] [doi: 10.2196/47884] [Medline: 37721792]
83. Parry M, Owadally T, O'Hara A, Nickerson N, Hart D. Community- and patient-partner engagement in women's cardiovascular disease research: a rapid review of the evidence. CJC Open 2024 Feb;6(2Part B):485-502 [FREE Full text] [doi: 10.1016/j.cjco.2023.12.016] [Medline: 38487065]
84. Alkureishi MA, Choo ZY, Rahman A, Ho K, Benning-Shorb J, Lenti G, et al. Digitally disconnected: qualitative study of patient perspectives on the digital divide and potential solutions. JMIR Hum Factors 2021 Dec 15;8(4):e33364 [FREE Full text] [doi: 10.2196/33364] [Medline: 34705664]
85. Hill JR, Brown JC, Campbell NL, Holden RJ. Usability-in-place-remote usability testing methods for homebound older adults: rapid literature review. JMIR Form Res 2021 Nov 02;5(11):e26181 [FREE Full text] [doi: 10.2196/26181] [Medline: 34726604]

Abbreviations

CAD: coronary artery disease
HRQoL: health-related quality of life
MRC: Medical Research Council
RCT: randomized controlled trial
REB: Research Ethics Board
REDCap: Research Electronic Data Capture
SMD: standardized mean difference
SUS: System Usability Scale

Edited by K Adapa; submitted 20.02.24; peer-reviewed by EM Noack, SY Fang, S Ania, V Quinten; comments to author 23.05.24; revised version received 10.12.24; accepted 31.03.25; published 17.04.25.

Please cite as:
Parry M, Huang T, Clarke H, Bjørnnes AK, Harvey P, Parente L, Norris C, Pilote L, Price J, Stinson JN, O'Hara A, Fernando M, Watt-Watson J, Nickerson N, Spiteri DeBonis V, Hart D, Faubert C
Development and Systematic Evaluation of a Progressive Web Application for Women With Cardiac Pain: Usability Study
JMIR Hum Factors 2025;12:e57583
URL: https://humanfactors.jmir.org/2025/1/e57583
doi:10.2196/57583
PMID:

Following Up Patients With Chronic Pain Using a Mobile App With a Support Center: Unicenter Prospective Study

Marta Antonia Gómez-González[1*], MD; Nicolas Cordero Tous[1*], MD, PhD; Javier De la Cruz Sabido[1*], MD; Carlos Sánchez Corral[1*], MD; Beatriz Lechuga Carrasco[1*], RNC; Marta López-Vicente[1*], MD; Gonzalo Olivares Granados[1,2*], MD, PhD

[1]Department of Neurosurgery, Hospital Universitario Virgen de las Nieves, Av. Juan Pablo II s/n, Granada, Spain

[2]Department of Human Anatomy, University of Granada, Granada, Spain

[*]all authors contributed equally

Corresponding Author:
Nicolas Cordero Tous, MD, PhD
Department of Neurosurgery, Hospital Universitario Virgen de las Nieves, Av. Juan Pablo II s/n, Granada, Spain

Abstract

Background: Chronic pain is among the most common conditions worldwide and requires a multidisciplinary treatment approach. Spinal cord stimulation is a possible treatment option for pain management; however, patients undergoing this intervention require close follow-up, which is not always feasible. eHealth apps offer opportunities for improved patient follow-up, although adherence to these apps tends to decrease over time, with rates dropping to approximately 60%. To improve adherence to remote follow-up, we developed a remote follow-up system consisting of a mobile app for patients, a website for health care professionals, and a remote support center.

Objective: Our objective was to evaluate patient adherence to remote follow-up using a system that includes a mobile app and a remote support center.

Methods: After review of the literature and approval of the design of the follow-up system by a multidisciplinary committee, a team of experts developed a system based on a mobile app, a website for health care professionals, and a remote support center. The system was developed in collaboration with health care professionals and uses validated scales to capture patients' clinical data at each stage of treatment (ie, pretreatment phase, trial phase, and implantation phase). Data were collected prospectively between January 2020 to August 2023, including the number of total surveys sent, surveys completed, SMS text message reminders sent, and reminder calls made.

Results: A total of 64 patients were included (n=40 women, 62.5%) in the study. By the end of the study, 19 (29.7%) patients remained in the pretreatment phase, 8 (12.5%) patients had completed the trial phase, and 37 (57.8%) reached the implantation phase. The mean follow-up period was 15.30 (SD 9.43) months. A total of 1574 surveys were sent, along with 488 SMS text message reminders and 53 reminder calls. The mean adherence rate decreased from 94.53% (SD 20.63%) during the pretreatment phase to 65.68% (SD 23.49%) in the implantation phase, with an overall mean adherence rate of 87.37% (SD 15.37%) for the app. ANOVA showed that adherence was significantly higher in the earlier phases of treatment (P<.001).

Conclusions: Our remote follow-up system, supported by a remote support center improves adherence to follow-up in later phases of treatment, although adherence decreased over time. Further studies are needed to investigate the relationship between adherence to the app and pain management.

(JMIR Hum Factors 2025;12:e60160) doi:10.2196/60160

KEYWORDS

pain management; mobile health; mHealth; eHealth; chronic pain; support center; mobile phone app; survey; follow-up; pain control; prospective study

Introduction

Chronic pain is one of the most common conditions globally and is associated with reduced quality of life, increased medical expenses, and significant economic costs [1]. Its prevalence ranges between 2% - 40% [1-3], and annual health care costs due to pain amount to US $300 billion in the United States [1] and €12 billion (US $12.3 billion) in Norway [2].

The treatment of chronic pain requires a multidisciplinary approach, with spinal cord stimulation being a potential treatment option. This technique is safe and effective [4] and has also shown benefits in management of other chronic pain conditions, such as complex regional pain syndromes, low back

pain, and other forms of pain. However, ensuring consistent follow-up for these patients is challenging, as they require close monitoring, which is not always feasible.

New technologies have opened up a lot of possibilities to address this challenge. The World Health Organization defines eHealth as the cost-effective and secure use of information and communication technologies to support health and health-related areas, including health services, health monitoring, health literature, health education, and knowledge and research [5].

There are numerous tools designed to facilitate medical care, including mobile apps, websites, and other platforms. While there is limited evidence to support their use, preliminary results are promising. For example, mobile apps used alongside conventional treatments have shown better results in the management of chronic pain than conventional interventions alone [6,7].

eHealth apps also offer an opportunity for closer patient follow-up, although adherence to treatment tends to decline over time, dropping to approximately 60%, as previous studies have shown [8,9]. To address this challenge, we have developed a system that includes a mobile app for patients, a website for health care professionals, and a remote support center.

The aim of this study is to assess patient adherence to follow-up care using a mobile app supported by a remote support center.

Methods

Overview

In 2019, researchers at the Department of Pain Surgery at Hospital Virgen de las Nieves in Granada, Spain, initiated approaches to improve the remote monitoring of patients implanted with epidural spinal cord stimulators for chronic pain management. The aim was to reduce the number of face-to-face consultations or hospital visits while providing additional support to these patients. Therefore, we decided to develop an app complemented by a support center to improve patient monitoring as much as possible. After reviewing the relevant literature, we adapted the tool kit published by Marvel et al [10] to suit our needs, following the four key steps outlined in Table 1. As we already had a multidisciplinary group, we sought the agreement of existing team members and assembled a group of experts. Instead of accelerating the project, we adapted our current workflows and protocols to the new working framework. We also developed the follow-up system after adapting the protocols already in place and enrolled patients only after the system was finalized.

Table . Project phases.

Johns Hopkins tool steps[a]	Project workflow (current study)
Define the problem and the digital tool	Define the problem and the digital tool
Creation of a multidisciplinary group	Creation of an expert team, approval by a multidisciplinary group
Seeking opportunities to accelerate the project	Adaptation of the existing clinical and educational protocol into the framework of the remote follow-up system
Involving professionals	Designing the follow-up system, considering the needs of the involved professionals, and subsequent enrollment of patients
Consulting different partners	Performing quality assessment
Conducting a clinical validation	Conducting a clinical validation

[a]Marvel et al [10].

The concept of a remote follow-up system was presented to a multidisciplinary committee composed of anesthesiologists, neurologists, neurosurgeons, neurophysiologists, and rehabilitation specialists. This committee meets monthly to assess patients with chronic pain who may benefit from interventional therapies such as spinal cord stimulation when conventional treatments have failed. The target population was identified as patients who could benefit from spinal cord stimulation and had not yet started the trial phase, to enable data collection before and after implantation.

The inclusion criteria for patients eligible for spinal cord stimulation system implantation at our center are shown in Textbox 1.

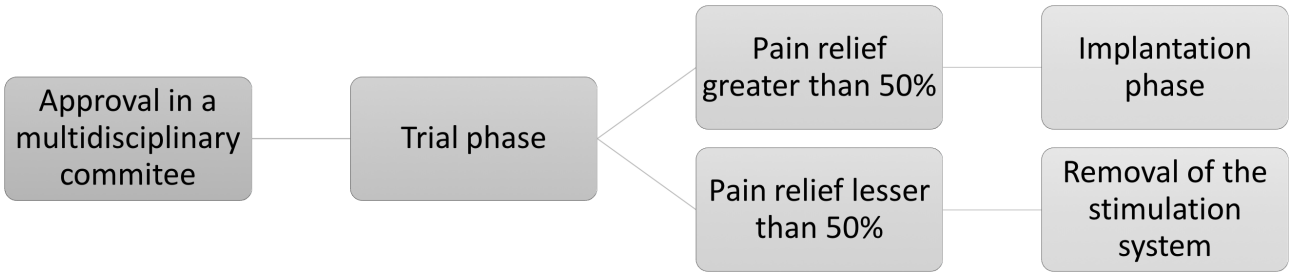
Textbox 1. Inclusion criteria.

- Chronic pain in a specific area of the body (patients experiencing diffuse pain were not accepted)
- Failure of conservative treatments, including rehabilitation and local infiltration
- A favorable psychological evaluation confirming that the patient can wear a spinal cord stimulation system
- No contraindications to surgery or allergies to system components
- No history or current substance use, especially opioids

Once the committee had approved the implantation of a spinal cord stimulator, the procedure was performed in 2 phases, as shown in Figure 1. In the trial phase, one or more electrodes were implanted in the epidural space and connected to an external stimulator. The efficacy of the system was then assessed over a 4-week period using the verbal numerical rating score

(VNRS). The intervention was considered effective if the patient's reported subjective perception of pain relief was at least 50%. If the trial phase was successful, it was followed by the implantation phase, in which a permanent stimulator was implanted in the adipose tissue of the abdomen or lower back. If the 50% threshold was not reached, the system was removed completely.

Figure 1. Algorithm used to evaluate the implantation of the spinal cord stimulation system.



Ethical Considerations

Ethics approval was obtained from the Comité de Investigación Provincial de Granada (study: NC-D-01; ethics committee reference: SICEIA-2020 - 000438), and informed consent for the procedure and follow-up system was obtained from all included patients. To maintain privacy, each patient was assigned a code with a number so that only health care professionals could access the personal data. The platform meets all the standards of Regulation (EU) 2016/679 of the European Parliament and of the Council of April 27, 2016, and has been approved by local and state authorities. This ensures that only health care professionals can access all the data collected and that data security is contractually guaranteed. The study participants have not been compensated in any way.

Design of the Follow-Up System

The remote follow-up system consists of 3 main components: a mobile app for patients, a website for health care professionals, and a remote follow-up center. The mobile app was designed to ensure close monitoring during all phases of treatment, using feedback from all health care professionals involved. The app was developed in collaboration with Clinical Care Connect (Persei) and customized by our team to meet our specific needs. It offers various features including educational content such as brochures and videos, questionnaires with validated scales, pop-up notifications, and a form to contact the remote support center. The validated scales used include Doleur Neuropathique 4 (DN4) to assess neuropathic pain, Oswestry Disability Index/Neck Disability Index (ODI/NDI) to assess back pain, VNRS to assess pain, and 36-item Short Form Survey (SF-36) to assess quality of life. In addition, patient-reported experience measures were used to assess experience and satisfaction with the app. Since pain is a subjective perception, the use of validated scales can help to measure it and compare it between patients.

The website designed by Persei (Vivarium S.L., Spain) can be accessed from anywhere with the necessary authorizations. It also provides alerts using different colors depending on the urgency of the detected problem: red for infection, yellow for insufficient pain relief, and blue for low battery. These alerts are monitored by the remote support center, which contacts the specialist by making a telephone call if necessary, such as, in the event of an infection. For nonemergency issues, such as insufficient pain relief, the support center contacts the patient to analyze the cause, and a priority appointment for treatment is scheduled. This remote support center was set up to increase the efficiency of treatment and improve pain management. The aim was to maximize treatment adherence and ensure the sustainability of the remote follow-up system. Previously, follow-up was carried out 1 month after implantation, 1 year after implantation, and subsequently, as required.

Data Collection and Analysis

We collected data from all patients who were approved for spinal cord stimulation between January 2020 and August 2023 and enrolled in the remote monitoring program. If the committee approved the patient for spinal cord stimulation, informed consent for the remote follow-up system was obtained in addition to consent for implantation. Data were collected automatically and prospectively through surveys completed via the app. The questionnaires were distributed according to the protocol shown in Textbox 2.

Textbox 2. Protocol used to send validated scales.

Validated scales used in the protocol used by the Virgen de las Nieves Hospital

- Pretreatment phase: Doleur Neuropathique 4 (DN4), verbal numerical rating score (VNRS), Oswestry Disability Index/Neck Disability Index (ODI/NDI), 36-item Short Form Survey (SF-36)
- Trial phase: Daily VNRS and patient-reported experience measures (PREMs) at the end of the trial phase
- Implantation phase: DN4, VNRS, ODI, and SF-36 after 1 and every 3 months; the PREMs survey was also distributed after 12 months

The support center sends a notification each time a survey needs to be completed, for a maximum of 3 times. If the questionnaires were not completed, a notification was sent to the hospital so that a specialized caregiver can contact the patient. One of the biggest challenges was ensuring privacy. To achieve this, each patient was coded with a number so that only health care professionals had access to the personal data and that data security was contractually guaranteed.

The data were collected automatically when a survey was answered via the app and stored in a private cloud. This allowed us to minimize potential errors. In addition, the support center guaranteed that the data collected via the app belonged to our hospital's patients and was only accessed and analyzed when a specialist at our center requested it. Demographic information was collected for data analysis. Adherence was measured by calculating the percentage of completed surveys relative to the total number of surveys sent. We analyzed adherence rate during each treatment phase and for the overall study, and compared adherence by demographic variables using ANOVA, adjusting for gender, age, diagnosis, and phase. Age was categorized into 2 groups, with 40 years as the cutoff point, since previous studies have found significant differences in prevalence rates for chronic pain and quality of life [11]. Regarding completeness, we only considered a survey as answered if it was completed. If data were missing, a notification was sent to the patient to complete the survey.

Results

Between January 2020 and August 2023, a total of 64 patients were enrolled in the study, comprising 24 (37.5%) men and 40 (62.5%) women. While identifying the cause of pain, 32 (50%) patients had complex regional pain syndromes, 18 (28%) had failed back syndrome, and 14 (22%) had other diagnoses such as phantom limb syndrome, adhesive arachnoiditis, coccydynia, and axonotmesis. Before the onset of pain, 19 (28%) patients had no fixed occupation, while 10 (18%) were involved in nonphysical and 35 (55%) in physical occupations such as farming or construction (Table 2).

Table . Summary of epidemiological data.

Variables	Patients (N=64), n (%)
Gender	
Men	24 (38)
Women	40 (63)
Age interval	
<40 years	9 (14)
≥40 years	55 (86)
Diagnoses	
CRPS[a]	32 (50)
FBS[b]	18 (22)
Others	14 (18)
Occupation	
Physical occupation	35 (55)
Nonphysical occupation	19 (18)
No fixed occupation	10 (28)
Phase	
Pretreatment	19 (30)
Trial	8 (13)
Implantation	37 (58)

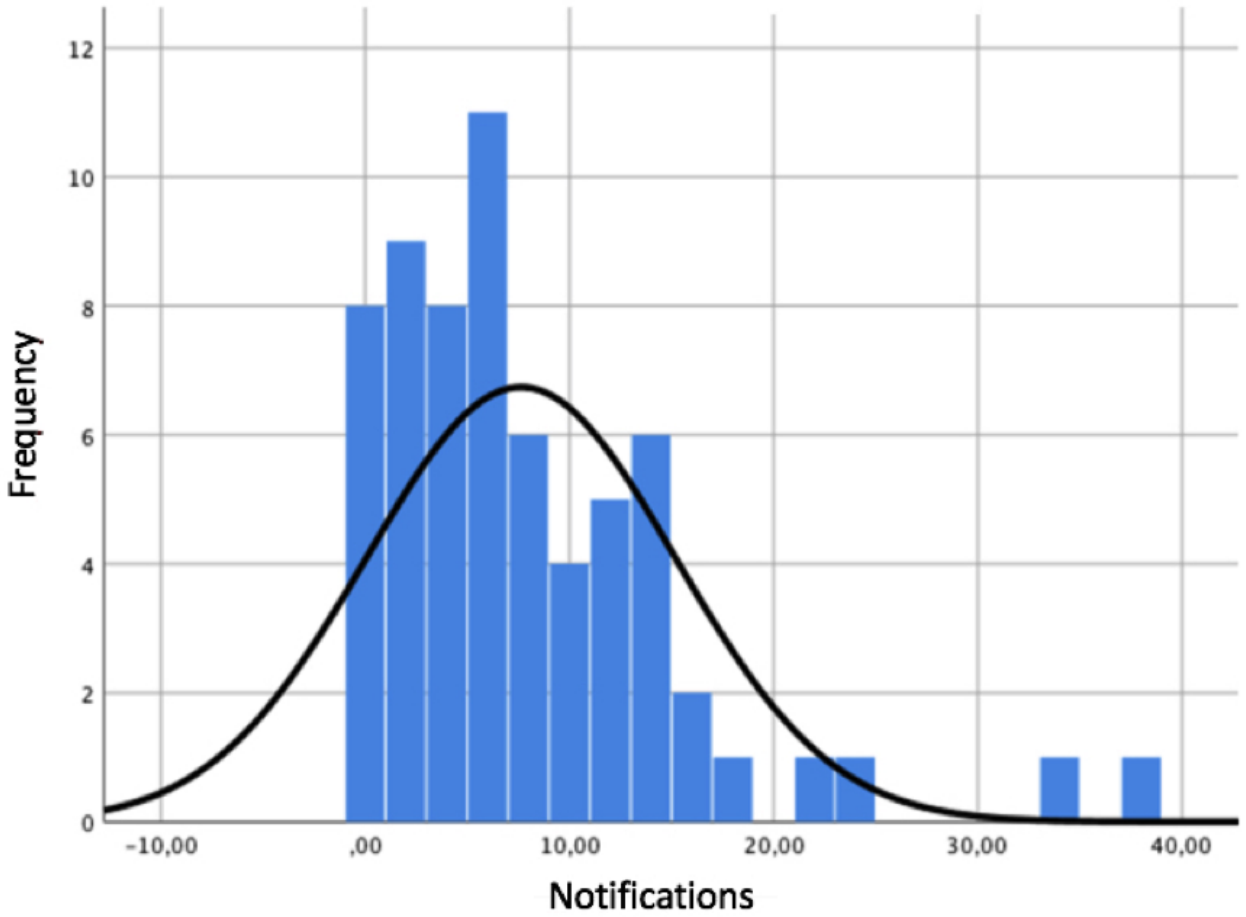
[a]CRPS: chronic regional pain syndrome.

[b]FBS: failed back syndrome.

At the end of the study, 19 (30%) patients were in the pretreatment group, awaiting the start of the trial phase. A total of 8 (13%) patients reached the trial phase but were not transferred to the implantation phase due to insufficient improvement in pain management. Further, a total of 37 (58%) patients advanced to the implantation phase (Table 2).

A total of 1586 surveys were distributed during the study. SMS text message notifications had to be sent to the patients 488 times, which corresponded to a mean of 7.62 (SD 7.57) SMS text messages per patient. However, 8 (13%) patients did not require SMS text messages notifications, while the remaining (n=56, 88%) patients required at least one notification (Figure 2). It was observed that several patients required more than 30 notifications. In addition, a total of 53 phone calls were made to remind patients to complete the surveys; however, 43 (67%) patients did not require any reminder calls.

Figure 2. Histogram with a normality curve showing the number of patients in groups (frequency) based on the number of required notifications (notifications).



The mean follow-up time was 15.38 (SD 9.43) months. A total of 12 (19%) patients were lost to follow-up, and 3 (4%) patients discontinued using the app (ie, 1 during the pretreatment phase and 2 patients during the implantation phase). For patients who discontinued using the app, any surveys not answered by them were considered failures.

Adherence was analyzed for each phase and overall, as shown in Figure 3, which represents the total number of surveys sent and answered in each phase, as well as the adherence rate in each phase. During the pretreatment phase, 256 surveys were sent, of which 242 were completed, corresponding to a mean adherence rate of 94.53% (SD 20.63%). During the trial phase, 645 surveys were sent, of which 499 were answered, corresponding to a mean adherence rate of 77.36% (SD 37.01%). During the implantation phase, 624 surveys were sent, of which 422 were answered, corresponding to a mean adherence of 65.68% (SD 23.49%). A total of 61 satisfaction surveys were sent (excluding 3 patients who discontinued using the app), of which 55 were answered, corresponding to a mean adherence rate of 92.11% (SD 18.47%). The overall mean adherence rate for using the app was 87.37% (SD 15.37%) across all phases (Table 3).

Figure 3. Surveys and adherence rate.

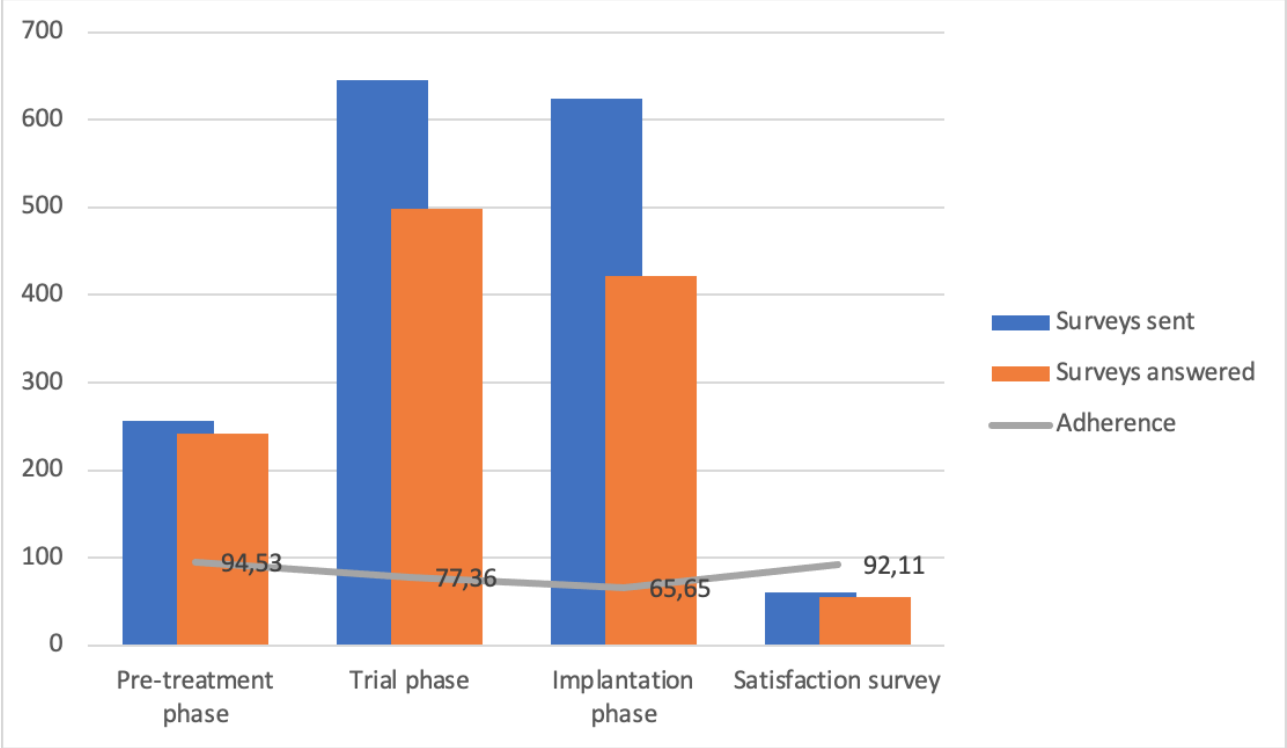


Table . Summary of surveys sent and answered and adherence in each phase.

Phase	Surveys sent, n[a]	Surveys answered, n[a]	Adherence rate[b] (%), mean (SD[c])
Pretreatment phase	256	242	94.53 (20.63)
Trial phase	645	499	77.36 (37.01)
Implantation phase	624	422	65.68 (23.49)
Satisfaction survey	61	55	92.11 (18.47)
Total adherence	1586	1218	87.37 (15.37)

[a]Number of surveys during each phase.

[b]Adherence rate was defined as the total surveys answered x 100/total surveys sent.

[c]The SD represents the variability of adherence rates among individual patients.

ANOVA revealed that adherence rates were significantly higher in the earlier phases of treatment (P<.001). There was also a trend indicating that women were more likely to complete satisfaction surveys (P=.17) and surveys in general (P=.64). However, women also required more SMS text message reminders (P=.70) and phone calls (P=.71). No significant differences in treatment adherence were found when comparing younger patients versus patients older than 40 years or when comparing adherence to treatment based on the cause of pain or type of occupation (Table 4).

Table . Summary of the ANOVA results.

Variables	Pretreatment phase adherence (%; n=64), mean (SD)	*P* value	Trial phase adherence (%; n=43), mean (SD)	*P* value	Implantation phase adherence (%; n=38), mean (SD)	*P* value	Satisfaction survey adherence (%; n=38), mean (SD)	*P* value	Global adherence (%), mean (SD)	*P* value
Gender		.81		.49		.18		.40		.80
Total	94.5 (20.6)		66.7 (23.5)		65.7 (23.5)		92.1 (18.5)		87.4 (15.4)	
Men	93.7 (22.4)		58.2 (31.3)		58.2 (31.3)		95.8 (14.4)		86.8 (19.1)	
Women	95.0 (19.8)		69.1 (18.6)		69.1 (18.6)		90.4 (20.1)		87.7 (12.9)	
Age (years)		.38		.29		.74		.90		.30
Total	94.5 (20.6)		77.36 (37.0)		65.7 (23.5)		92.1 (18.5)		83.4 (15.4)	
≤40	88.9 (33.3)		62.9 (42.2)		68.4 (19.9)		92.9 (18.9)		82.5 (14.9)	
>40	95.5 (18.1)		80.2 (34.6)		65.1 (25.6)		91.3 (18.7)		88.2 (15.4)	
Diagnoses		.42		.63		.57		.83		.35
Total	94.5 (10.6)		77.4 (37.0)		65.7 (23.5)		92.1 (18.5)		87.4 (15.6)	
CRPS[a]	100 (0)		72.8 (42.2)		63.1 (29.9)		92.8 (17.9)		85.7 (16.9)	
FBS[b]	94.5 (21.8)		81.7 (30.2)		73.5 (12.8)		93.8 (17.7)		92.7 (9.7)	
Others	90.3 (25.9)		85.3 (28.3)		64.8 (9.7)		88.9 (22.0)		86.2 (15.9)	
Phase		.61		.51		.60		.67		.001
Total	94.5 (20.6)		—[c]		—		—		87.4 (15.4)	
Pretreatment	96.1 (12.5)		77.8 (21.1)		77.8 (0)		100 (0)		96.1 (12.5)	
Trial	100 (0)		65.4 (4)		65.4 (23.8)		91.9 (18.7)		94.3 (9.2)	
Implantation	92.6 (25.6)		77.4 (37.0)		65.7 (23.5)		92.1 (18.5)		81.4 (15.2)	

[a]CRPS: complex regional pain syndromes.

[b]FBS: fail back syndrome.

[c]Not applicable.

Discussion

Comparison With Previous Research

In recent years, there has been increasing interest in the use of pain control apps for various purposes, such as postsurgical follow-up, pain self-management, and chronic pain management [6-10,12-14]. However, many of these tools have not been as effective as expected, often due to inadequate planning and a lack of preliminary studies to anticipate necessary roles [15]. However, some studies have demonstrated that these apps can outperform traditional follow-up methods [6,14]. Our study introduces a remote support center, a feature that we believe is not yet widely adopted internationally. The aim of this study is to determine whether the inclusion of a support center improves adherence compared to other systems.

An analysis of the medical apps across various app stores shows that approximately 86% of these apps were developed without the involvement of medical professionals [16]. Lalloo et al [17] examined 1019 apps developed for postoperative pain management or education and found that only 10 apps met the established criteria. When evaluating apps developed with the involvement of health care professionals, only 5 (0.49%) of them were deemed suitable. Similarly, Bhattarai et al [16] examined 373 apps focused on arthritis pain self-management with only 4 apps meeting the Stanford criteria for pain self-management. Portelli and Eldred [18] evaluated 195 apps for pain management against the guidelines for cognitive behavioral therapy. Of these, only 6 (3%) of these apps met the standards, leading to the conclusion that neither health care professionals nor patients were involved in their development. Despite these limitations, a recent meta-analysis of 4767 patients in 22 randomized trials found that these apps offer a small but significant improvement in long-term pain management [19].

For patients undergoing spinal cord stimulation, close follow-up is required to adjust the type and parameters of stimulation, which usually necessitates face-to-face visits. This increases the time and cost burdens for patients and caregivers, especially for those who live far from specialized centers [20]. Our remote follow-up system was developed to overcome these challenges and ensure optimal patient monitoring. The mobile app in combination with the remote support center reduces the need

for in-person examinations while ensuring a high standard of care. The app's alert system helps to detect problems at an early stage, such as infections or low battery levels so that immediate action can be taken.

In addition, the use of validated scales during an in-person consultation is challenging due to the limited time available. Consequently, this system has allowed a better understanding of the patient's condition. Furthermore, some patients have found that expressing themselves via the app helps them [20] in experiencing privacy and without any time pressure.

Adherence to medical app use is crucial as studies show that 75% of users stop using an app within 48 hours of downloading, and 15% delete it after initial use [21]. A review by Wikström et al [9] found that none of the studies focused on improving adherence or motivating users to continue using medical apps. The easiest way to increase motivation is through notifications, which is why our study focused on the role of the remote support center. We consider this as a key factor in the success of our study. However, numerous motivational elements in medical apps have been analyzed in previous studies, as summarized in Table 5.

Table . Motivational elements in eHealth apps.

Studies	Motivational elements								
	Tablet lent to the patient	Chat	Involvement of relatives	Educational texts, pictures, and/or videos	Follow-up of auto evaluation and graphic results	Personalized follow-up	Sharing media elements	Notifications and reminders	Alerts
Alam et al [20]		✓[a]		✓				✓	
Pecorelli et al [22]	✓			✓	✓			✓	
Perdoncini et al [23]		✓		✓					
Shah et al [24]		✓		✓					
Davidovitch et al [25]		✓		✓			✓		
Felbaum et al [26]		✓		✓			✓	✓	
Glauser et al [27]		✓		✓	✓		✓	✓	
Gustavell et al [28]		✓		✓	✓			✓	✓
Hou et al [29]		✓		✓		✓		✓	✓
Pickens et al [30]		✓	✓	✓				✓	
Timmers et al [31]				✓	✓		✓	✓	
van der Meij et al [32]				✓		✓			
Mundi et al [33]				✓				✓	
This study		✓	✓	✓	✓	✓	✓	✓	✓

[a]Element analyzed in the corresponding study.

Our remote follow-up system was developed based on relevant literature to maximize adherence and achieved an overall mean adherence rate of 87.37% (SD 15.37%). Comparison of this rate with previous studies is difficult, as most of these focused on adherence to a treatment or intervention rather than the use of an app. For example, Gomis-Pastor et al [34] showed that the use of an app significantly improved adherence to treatment and medication regimen, which led to significantly better symptom control, as found in a 2022 review [35]. However, Wikström et al [9] have shown that adherence to app use decreases over time, potentially due to improvement in health status or loss of interest [29], among other factors [26,29,32]. Our study aligns with these findings and shows a decrease in adherence from 94.53% in the pretreatment phase to 77.36% in the testing phase and 65.58% in the implantation phase [36-38].

Limitations

The main limitation of our study is the specificity of the target population, as spinal cord stimulation is a specialized treatment for patients experiencing chronic pain. These patients often display a significant emotional component that can influence the outcomes of the interventions, posing challenges during designing this study and its remote follow-up system. In addition, these systems rely on new technologies that can be challenging for older patients, potentially limiting their ability to fully benefit from them. By using self-reported data, we also acknowledged that omissions or misunderstandings in the surveys may be introduced, which we attempted to minimize using validated scales. Finally, setting up a system similar to this study requires a significant financial investment.

Conclusions

Digital systems are a part of medical care, and it is important that health care professionals are involved in the development of tools to ensure the achievement of desired standards. We have developed a remote monitoring system for patients undergoing spinal cord stimulation, based on scientific evidence and supported by a remote support center that improves treatment adherence. This system also provides us with detailed information about each patient. However, there is a tendency for patients to abandon the app usage over time, which could be related to long-term pain control. Further studies are needed to investigate the relationship between adherence and pain control.

Conflicts of Interest

We have received financial support from Medtronic IHS CareConnect Pain to cover publication costs for this article. However, we declare no conflicts of interest as we implant spinal cord stimulation systems from different brands and treat all equally. In addition, none of the authors work for Medtronic IHS CareConnect Pain or receive economic compensation from Medtronic IHS CareConnect Pain.

References

1. Gaskin DJ, Richard P. The economic costs of pain in the United States. J Pain 2012 Aug;13(8):715-724. [doi: 10.1016/j.jpain.2012.03.009] [Medline: 22607834]
2. Stubhaug A, Hansen JL, Hallberg S, Gustavsson A, Eggen AE, Nielsen CS. The costs of chronic pain–Long-term estimates. Eur J Pain 2024 Jul;28(6):960-977. [doi: 10.1002/ejp.2234] [Medline: 38214661]
3. Yong RJ, Mullins PM, Bhattacharyya N. Prevalence of chronic pain among adults in the United States. Pain 2022 Feb 1;163(2):e328-e332. [doi: 10.1097/j.pain.0000000000002291] [Medline: 33990113]
4. Kapural L, Peterson E, Provenzano DA, Staats P. Clinical evidence for spinal cord stimulation for failed back surgery syndrome (FBSS): systematic review. Spine (Phila Pa 1976) 2017 Jul 15;42 Suppl 14:S61-S66. [doi: 10.1097/BRS.0000000000002213] [Medline: 28441313]
5. eHealth. World Health Organization - Regional Office for the Eastern Mediterranean. URL: https://www.emro.who.int/health-topics/ehealth/ [accessed 2024-04-03]
6. Toelle TR, Utpadel-Fischler DA, Haas KK, Priebe JA. App-based multidisciplinary back pain treatment versus combined physiotherapy plus online education: a randomized controlled trial. NPJ Digit Med 2019;2:34. [doi: 10.1038/s41746-019-0109-x] [Medline: 31304380]
7. Rughani G, Nilsen TIL, Wood K, et al. The selfBACK artificial intelligence-based smartphone app can improve low back pain outcome even in patients with high levels of depression or stress. Eur J Pain 2023 May;27(5):568-579. [doi: 10.1002/ejp.2080] [Medline: 36680381]
8. Kerckhove N, Delage N, Cambier S, et al. eDOL mHealth app and web platform for self-monitoring and medical follow-up of patients with chronic pain: observational feasibility study. JMIR Form Res 2022 Mar 2;6(3):e30052. [doi: 10.2196/30052] [Medline: 35234654]
9. Wikström L, Schildmeijer K, Nylander EM, Eriksson K. Patients' and providers' perspectives on e-health applications designed for self-care in association with surgery - a scoping review. BMC Health Serv Res 2022 Mar 23;22(1):386. [doi: 10.1186/s12913-022-07718-8] [Medline: 35321707]
10. Marvel FA, Wang J, Martin SS. Digital health innovation: a toolkit to navigate from concept to clinical testing. JMIR Cardio 2018 Jan 18;2(1):e2. [doi: 10.2196/cardio.7586] [Medline: 31758761]
11. Rustøen T, Wahl AK, Hanestad BR, Lerdal A, Paul S, Miaskowski C. Age and the experience of chronic pain: differences in health and quality of life among younger, middle-aged, and older adults. Clin J Pain 2005;21(6):513-523. [doi: 10.1097/01.ajp.0000146217.31780.ef] [Medline: 16215337]
12. Jaensson M, Dahlberg K, Eriksson M, Grönlund Å, Nilsson U. The development of the recovery assessments by phone points (RAPP): a mobile phone app for postoperative recovery monitoring and assessment. JMIR Mhealth Uhealth 2015 Sep 11;3(3):e86. [doi: 10.2196/mhealth.4649] [Medline: 26362403]
13. Bhatia A, Kara J, Janmohamed T, et al. User engagement and clinical impact of the manage my pain app in patients with chronic pain: a real-world, multi-site trial. JMIR Mhealth Uhealth 2021 Mar 4;9(3):e26528. [doi: 10.2196/26528] [Medline: 33661130]

14. Priebe JA, Haas KK, Moreno Sanchez LF, et al. Digital treatment of back pain versus standard of care: the cluster-randomized controlled trial, Rise-uP. J Pain Res 2020;13:1823-1838. [doi: 10.2147/JPR.S260761] [Medline: 32765057]
15. Salazar A, de Sola H, Failde I, Moral-Munoz JA. Measuring the quality of mobile apps for the management of pain: systematic search and evaluation using the mobile app rating scale. JMIR Mhealth Uhealth 2018 Oct 25;6(10):e10718. [doi: 10.2196/10718] [Medline: 30361196]
16. Bhattarai P, Newton-John TRO, Phillips JL. Quality and usability of arthritic pain self-management apps for older adults: a systematic review. Pain Med 2018 Mar 1;19(3):471-484. [doi: 10.1093/pm/pnx090] [Medline: 28541464]
17. Lalloo C, Shah U, Birnie KA, et al. Commercially available smartphone apps to support postoperative pain self-management: scoping review. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e162. [doi: 10.2196/mhealth.8230] [Medline: 29061558]
18. Portelli P, Eldred C. A quality review of smartphone applications for the management of pain. Br J Pain 2016 Aug;10(3):135-140. [doi: 10.1177/2049463716638700] [Medline: 27583140]
19. Rammo R, Gostkowski M, Rasmussen PA, Nagel S, Machado A. The need for digital health solutions in deep brain stimulation for Parkinson's disease in the time of COVID-19 and beyond. Neuromodulation 2021 Feb;24(2):331-336. [doi: 10.1111/ner.13307] [Medline: 33174292]
20. Alam M, Banwell C, Olsen A, Lokuge K. Patients' and doctors' perceptions of a mobile phone-based consultation service for maternal, neonatal, and infant health care in Bangladesh: a mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11842. [doi: 10.2196/11842] [Medline: 31008716]
21. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97. [doi: 10.2196/jmir.4883] [Medline: 27185295]
22. Pecorelli N, Fiore JF Jr, Kaneva P, et al. An app for patient education and self-audit within an enhanced recovery program for bowel surgery: a pilot study assessing validity and usability. Surg Endosc 2018 May;32(5):2263-2273. [doi: 10.1007/s00464-017-5920-3] [Medline: 29098431]
23. Perdoncini NN, Schussel JL, Amenábar JM, Torres-Pereira CC. Use of smartphone video calls in the diagnosis of oral lesions: teleconsultations between a specialist and patients assisted by a general dentist. J Am Dent Assoc 2021 Feb;152(2):127-135. [doi: 10.1016/j.adaj.2020.10.013] [Medline: 33494867]
24. Shah SS, Gvozdanovic A, Knight M, Gagnon J. Mobile app-based remote patient monitoring in acute medical conditions: prospective feasibility study exploring digital health solutions on clinical workload during the COVID crisis. JMIR Form Res 2021 Jan 15;5(1):e23190. [doi: 10.2196/23190] [Medline: 33400675]
25. Davidovitch RI, Anoushiravani AA, Feng JE, et al. Home health services are not required for select total hip arthroplasty candidates: assessment and supplementation with an electronic recovery application. J Arthroplasty 2018 Jul;33(7S):S49-S55. [doi: 10.1016/j.arth.2018.02.048] [Medline: 29588123]
26. Felbaum DR, Stewart JJ, Anaizi AN, Sandhu FA, Nair MN, Voyadzis JM. Implementation and evaluation of a smartphone application for the perioperative care of neurosurgery patients at an academic medical center: implications for patient satisfaction, surgery cancelations, and readmissions. Oper Neurosurg (Hagerstown) 2018 Mar 1;14(3):303-311. [doi: 10.1093/ons/opx112] [Medline: 28541569]
27. Glauser G, Ali ZS, Gardiner D, et al. Assessing the utility of an IoS application in the perioperative care of spine surgery patients: the NeuroPath pilot study. Mhealth 2019;5(40):40. [doi: 10.21037/mhealth.2019.09.01] [Medline: 31620467]
28. Gustavell T, Sundberg K, Segersvärd R, Wengström Y, Langius-Eklöf A. Decreased symptom burden following surgery due to support from an interactive app for symptom management for patients with pancreatic and periampullary cancer. Acta Oncol 2019 Sep;58(9):1307-1314. [doi: 10.1080/0284186X.2019.1633473] [Medline: 31284797]
29. Hou J, Yang R, Yang Y, et al. The effectiveness and safety of utilizing mobile phone-based programs for rehabilitation after lumbar spinal surgery: multicenter, prospective randomized controlled trial. JMIR Mhealth Uhealth 2019 Feb 20;7(2):e10201. [doi: 10.2196/10201] [Medline: 30785406]
30. Pickens R, Cochran A, Tezber K, et al. Using a mobile application for real-time collection of patient-reported outcomes in hepatopancreatobiliary surgery within an ERAS® pathway. Am Surg 2019 Aug 1;85(8):909-917. [doi: 10.1177/000313481908500847] [Medline: 32051068]
31. Timmers T, Janssen L, van der Weegen W, et al. The effect of an app for day-to-day postoperative care education on patients with total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2019 Oct 21;7(10):e15323. [doi: 10.2196/15323] [Medline: 31638594]
32. van der Meij E, Anema JR, Leclercq WKG, et al. Personalised perioperative care by e-health after intermediate-grade abdominal surgery: a multicentre, single-blind, randomised, placebo-controlled trial. Lancet 2018 Jul;392(10141):51-59. [doi: 10.1016/S0140-6736(18)31113-9]
33. Mundi MS, Lorentz PA, Grothe K, Kellogg TA, Collazo-Clavell ML. Feasibility of smartphone-based education modules and ecological momentary assessment/intervention in pre-bariatric surgery patients. Obes Surg 2015 Oct;25(10):1875-1881. [doi: 10.1007/s11695-015-1617-7] [Medline: 25702141]
34. Gomis-Pastor M, Mirabet Perez S, Roig Minguell E, et al. Mobile health to improve adherence and patient experience in heart transplantation recipients: the mHeart trial. Healthcare (Basel) 2021 Apr 14;9(4):463. [doi: 10.3390/healthcare9040463] [Medline: 33919899]

35. Collado-Borrell R, Escudero-Vilaplana V, Narrillos-Moraza Á, Villanueva-Bueno C, Herranz-Alonso A, Sanjurjo-Sáez M. Patient-reported outcomes and mobile applications. a review of their impact on patients' health outcomes. Farm Hosp 2022 Apr 27;46(3):173-181. [Medline: 36183211]
36. Wang TF, Huang RC, Yang SC, Chou C, Chen LC. Evaluating the effects of a mobile health app on reducing patient care needs and improving quality of life after oral cancer surgery: quasiexperimental study. JMIR Mhealth Uhealth 2020 Jul 27;8(7):e18132. [doi: 10.2196/18132] [Medline: 32716303]
37. Higgins J, Chang J, Hoit G, Chahal J, Dwyer T, Theodoropoulos J. Conventional follow-up versus mobile application home monitoring for postoperative anterior cruciate ligament reconstruction patients: a randomized controlled trial. Arthroscopy 2020 Jul;36(7):1906-1916. [doi: 10.1016/j.arthro.2020.02.045] [Medline: 32268161]
38. Armstrong KA, Coyte PC, Brown M, Beber B, Semple JL. Effect of home monitoring via mobile app on the number of in-person visits following ambulatory surgery: a randomized clinical trial. JAMA Surg 2017 Jul 1;152(7):622-627. [doi: 10.1001/jamasurg.2017.0111] [Medline: 28329223]

Abbreviations

DN4: Doleur Neuropathique 4
ODI/NDI: Oswestry Disability Index/Neck Disability Index
SF-36: 36-item Short Form Survey
VNRS: verbal numeral rating score

Edited by C Jacob; submitted 05.05.24; peer-reviewed by AM Pinard, Y Chu; revised version received 26.11.24; accepted 01.12.24; published 22.01.25.

Please cite as:
Gómez-González MA, Cordero Tous N, De la Cruz Sabido J, Sánchez Corral C, Lechuga Carrasco B, López-Vicente M, Olivares Granados G
Following Up Patients With Chronic Pain Using a Mobile App With a Support Center: Unicenter Prospective Study
JMIR Hum Factors 2025;12:e60160
URL: https://humanfactors.jmir.org/2025/1/e60160
doi:10.2196/60160

Original Paper

Exploring Older Adults' Needs for a Healthy Life and eHealth: Qualitative Interview Study

Paula Valkonen[1], MA (Art and Des); Sari Kujala[1*], PhD; Kaisa Savolainen[1*], MSc (Tech), DA (Art and Des); Riina-Riitta Helminen[2,3], MD, LLLic, DSc (Tech)

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Suomen Terveystalo Oy, Suomen Terveystalo Oy, Helsinki, Finland

[3]Department of Industrial Engineering and Management, Aalto University, Espoo, Finland

[*]these authors contributed equally

Corresponding Author:
Paula Valkonen, MA (Art and Des)
Department of Computer Science
Aalto University
P.O.Box 15400
Espoo, FI-00076
Finland
Phone: 358 503862768
Email: paula.valkonen@aalto.fi

Abstract

Background: Aging brings physical and life changes that could benefit from eHealth services. eHealth holistically combines technology, tasks, individuals, and contexts, and all these intertwined elements should be considered in eHealth development. As users' needs change with life situations, including aging and retirement, it is important to identify these needs at different life stages to develop eHealth services for well-being and active, healthy lives.

Objective: This study aimed to (1) understand older adults' everyday lives in terms of well-being and health, (2) investigate older adults' needs for eHealth services, and (3) create design recommendations based on the findings.

Methods: A total of 20 older adults from 2 age groups (55 to 74 years: n=12, 60%; >75 years: n=8, 40%) participated in this qualitative interview study. The data were collected remotely using a cultural probes package that included diary-based tasks, sentence completion tasks, and 4 background questionnaires; we also performed remote, semistructured interviews. The data were gathered between the fall of 2020 and the spring of 2021 in Finland as a part of the Toward a Socially Inclusive Digital Society: Transforming Service Culture (DigiIN) project (2019 to 2025).

Results: In the daily lives of older adults, home-based activities, such as exercising (72/622, 11.6% of mentions), sleeping (51/622, 8.2% of mentions), and dining and cooking (96/622, 15.4% of mentions), promoted well-being and health. When discussing their needs for eHealth services, participants highlighted a preference for a chat function. However, they frequently mentioned barriers and concerns such as the lack of human contact, inefficiency, and difficulties using eHealth systems. Older adults value flexibility; testing possibilities (eg, trial versions); support for digital services; and relevant, empathetically offered content with eHealth services on short-term and long-term bases in their changing life situations.

Conclusions: Many older adults value healthy routines and time spent at home. The diversity of older adults' needs should be considered by making it possible for them to manage their health safely and flexibly on different devices and channels. eHealth services should adapt to older adults' life changes through motivation, personalized content, and appropriate functions. Importantly, older adults should still have the option to not use eHealth services.

(JMIR Hum Factors 2025;12:e50329) doi:10.2196/50329

KEYWORDS

older adults; eHealth; needs; retirement; well-being; cultural probes; sentence completion; human-computer interaction

Introduction

Background

Aging involves physical and life changes [1-4], where eHealth services have the potential to provide support and benefits [4-7]. The prevalence of many diseases increases with age [8,9], and multimorbidity is common among people aged ≥65 years [10,11]. Conversely, healthy behaviors can extend people's lives [12-14].

eHealth can help people nurture their health and well-being, potentially extending their lives. eHealth can also make treatment more accessible, promote treatment continuity, enhance communication, help shared decision-making, and enable patient self-management [10]. Older adults seem to be generally positive toward eHealth in a health management context [15]. They can benefit from eHealth services, as they often have complex health issues and engage in many self-management tasks supporting their health and well-being [1,5,16].

However, older adults sometimes face challenges with eHealth services [5,16,17], including a lack of skills and interest [3,16,18]. Hirvonen et al [19] reported that some older adults find the content or functionalities of eHealth services irrelevant (eg, [20,21]). Sometimes, older adults might even feel that digital services are not "meant for them" (eg, [20,22]). Thus, older adults might have a risk of exclusion from digital services [16,18,23,24] and society [16,23].

Older adults have not always been involved in technological development, especially in mainstream technology [25-27]. At the same time, the older population is growing [28], and health care costs are rising [29]. Stephanidis et al [30] encourage understanding emerging new technology possibilities and context-based user needs when designing digital solutions for the health and well-being (hereafter, *eHealth services* or *eHealth*) of older adults and a better understanding of the role of technology in older adults' lives [30]. Ideally, the technology should be harmonized with these adults' lives [15].

European older adults regularly engage in work, household chores, exercise, cultural activities, and tourism [31]. On the basis of the literature, many older adults in Europe live in rural areas, where they have easy access to nature and can engage in refreshing activities [31]. On the other hand, older adults with urban lifestyles access a variety of services [31]. For example, in Finland and Sweden, rural areas are in the north and distant from health care services, and eHealth services offer a promising option for taking care of one's health [32]. Physical activities might support older adults' health situation during aging. In 2017, in total 43.2% of Europeans aged 50 to 64 years and 44.5% of those aged 65 to 74 years old spent >3 hours per week engaging in physical activity [31]. On the other hand, older adults' daily routines differ. Older adults are a heterogeneous group of people, and aging is a process, not a static state [33]. Therefore, it is important to obtain qualitative insights into older adults' activities. These could help inform the idea and role of eHealth services and harmonize it with older adults' routines [15,30]. We know that older adults use eHealth services, for example, depending on their digital devices, capabilities, interests, or education [31,34]. Statistically, 100% of Europeans have, at least in theory, access to eHealth services or health information [35].

The best approach to eHealth service development means selecting the optimal approach case by case [36]. In eHealth, technology, tasks, individuals, and contexts are holistically combined, which should all be considered in developments [36-39]. eHealth services should genuinely support people pursuing healthier lifestyles and be accessible to all [10], although older adults may perceive the technology differently from other generations [1,40]. Humanity is needed alongside IT. Therefore, design decisions should also help ensure that as a part of eHealth services development, the human touch of personal care will not be removed [30,41].

Mapping user needs is important because without understanding needs, designing high-quality solutions is challenging [42]. User needs can be defined as the problems that prevent users from reaching their goals or possibilities to support them in reaching these goals [43]. User needs can be personal or more social in nature, and they can be related, for example, to available information or technology functions [44]. Poor understanding of users, user needs, and use contexts increases the risk of failure in eHealth service development [39]. On the other hand, user needs are not a stable phenomenon; they might change with age or the use of technology [45]. However, not much is known about what kind of user needs for eHealth services arise in older adults' lives, which is unfortunate because understanding user needs is a great innovation source [46]. In addition, Hirvonen et al [19] recommend further investigation into older adults' eHealth service use as a part of their daily lives.

Goal of This Study

We aimed to understand older adults' everyday lives in terms of well-being and health and their needs for eHealth services. We investigated 2 life stages: working at an older age and life after retirement, possibly with a chronic disease. We were interested in the health and well-being practices of 2 age groups to understand older adults' needs for eHealth services. Investigating these aspects allows developers and designers to better understand how eHealth services can meet older adults' needs and support their healthy living.

Users' needs change with life situations, including aging [47]. Therefore, if we want to harmonize eHealth services with older adults' everyday use, as is recommended by Cabrita et al [15], and to support technology development for well-being and an active, healthy lifestyle, it is important to identify user needs at different stages of life and daily routines in different life situations. The research questions were as follows:

1. What is everyday life like for older adults?
2. What kind of needs do older adults have for eHealth services?

Prior Work

Qualitative Methods for Investigation of Needs

User needs can be explored in many ways. In the human-computer interaction field [48,49], designers and

researchers have several methods for investigating user needs. Soraghan et al [50] and Dickinson et al [51] encourage researchers and designers to step into older adults' homes to assess technology use and empathize with respondents' lives, for example, through in-home interviews and observations. Understanding of older adults can be collected through remote, semistructured interviews [52]. Projective techniques, such as cultural probes [53-57] or sentence completion tasks [58], in which the participants provide information about themselves (instead of the researcher asking them direct questions), can be fruitful methods for gaining a deeper understanding of end users when the topic is challenging to verbalize or includes sensitive aspects [42]. For example, the sentence completion technique can be used to gather both an understanding of end users' values and needs as well as inspirational data for design and discussion activation [42,59,60].

Cultural probes are designed to trigger older adults to provide inspirational information about their lives to designers [53]. Following development, they have been widely used for various purposes, including collecting information about caregivers [56], investigating patient experiences with an eHealth service [54], and in other health care contexts [55,57]. Cultural probes are packages or toolkits of various tasks that end users complete and document [53]. While cultural probes show potential as a projective technique for gathering user needs [42], they appear to be underused in identifying older adults' needs for eHealth services.

Older Adults' Needs and Attitudes Toward Technology

Older adults' needs and attitudes toward technology vary. For example, some older adults do not trust computers and find the terminology and content of computers and digital services confusing [17,18,61]. They might even fear computers [61]. Their self-confidence with computers is sometimes fragile, and usability issues can be disastrous for their self-confidence [61]. In addition, older adults as a user group are much more diverse than many traditional user groups; they may have problems such as sensory loss, challenges with language, and attitudes to technology [61].

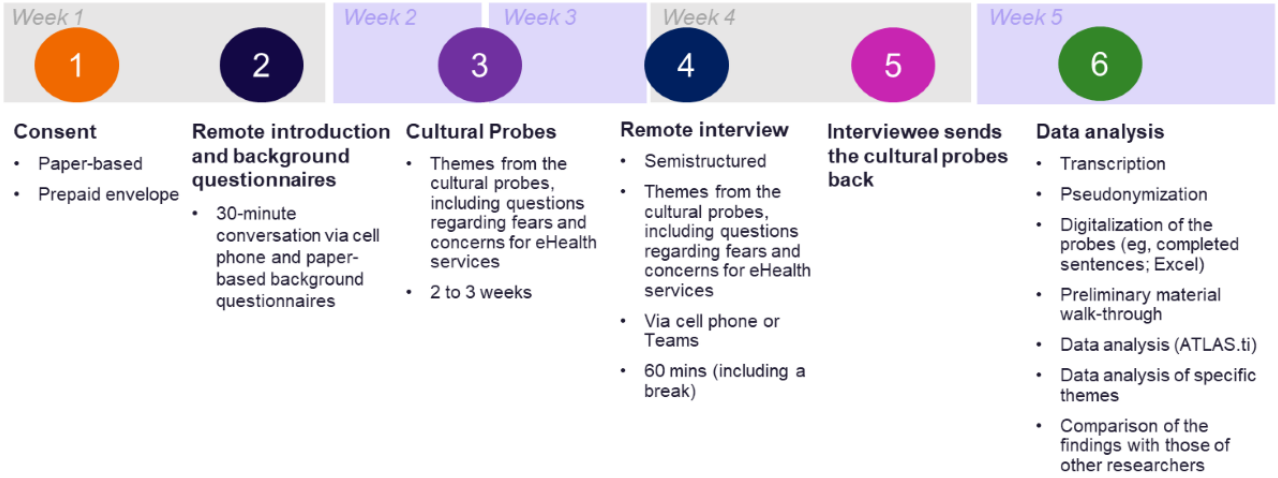
Many older adults appreciate easy access to eHealth services and their medical records, which should be offered in an easy, understandable, and secure way [5,52]. When medical information has compact and clearly labeled contents in well-organized menus [62], clear interface language, and linear navigation [20], this can assist older adults. Backonja et al [47] recommend eHealth services that adjust functionality and content based on the user's needs. Similarly, Cabrita et al [15] recommend personalization of design and functionality, thus lowering older end users' fears toward technology by allowing it to support its users and even offer empathy and sympathy to them.

Methods

Overview

This empirical qualitative study consisted of cultural probes [53,63], which included 4 background information questionnaires (*background questionnaires*), sentence completion tasks [42,58,59], and a diary-based exercise [64-66], and remote interviews [52]. Data collection included 6 phases, which all played an important role (Figure 1). Data collection started with filling out a consent form. After that, background information was collected, and a short remote interview was conducted. Then, the participant had time to respond to the cultural probe package (ie, sentence completion tasks and diary-based exercise). Participation ended with a long remote interview and the return of the probes package to the researcher. Finally, the data were analyzed.

Figure 1. Research procedure.



Sampling and Recruitment

The research was conducted in Finland between the fall of 2020 and the spring of 2021 (ie, during the COVID-19 pandemic) as part of the DigiIN project (2019 to 2025).

Defining an older adult purely by age is somewhat complex [67]. Generally, they are people in late adulthood. In line with the study by Ware et al [5], this study focused on older age groups associated with a higher risk of chronic disease, with participants from 2 age groups (those aged between 55 and 74

years and employed and those aged >75 years and retired). Group 1 participants (ie, those who were still working) were relevant because aging includes preparing for retirement [68].

The participants were selected through purposive sampling (digital survey) and snowball sampling [69]. Group 1 participants (12/20, 60%) were recruited via a digital survey [70]. Group 2 participants (8/20, 40%) were recruited in 2 ways: via the digital survey and snowball sampling. To ensure a diverse and engaged sample for this long-term study, participants' interest in digital services was assessed in advance. The groups were formed according to the life situation: group 1 participants were still working, and group 2 participants were retired. The age category of group 1 was between 55 and 74 years (later, 55-74 years). The age category of group 2 was between 75 and >90 years (later, >75 years). Participants in group 1 had different professions, such as yoga teacher, director, project manager, or entrepreneur. The participants were from different parts of Finland but did not fully cover the whole country.

After participants expressed potential interest in participating in the study, the researcher explained the study in a telephone call. This call lasted for 15 to 30 minutes, depending on the participant's questions. If the participant was still interested, the researcher mailed or emailed the information letter that described this study in detail and a consent form to the participants for their signature.

Ethical Considerations

The empirical study was conducted as part of the DigiIN project (2019 to 2025). The study protocol was reviewed and approved by the Ethical Review Board of Aalto University (95_03.04_2019_DigiIN). All participants provided their voluntary, informed, and written consent. Patients' ability and willingness to participate in the study were confirmed through a phone conversation before they signed the consent form. During the call, the study process was explained, and it was emphasized that participation was voluntary, with the option to withdraw at any time. This was also mentioned in the written consent form.

Research Procedure

The study comprised six phases (Figure 1):

1. First, the study purpose and research procedure were introduced.
2. Second, the cultural probe packages, including 4 background questionnaires, were delivered. The study procedure and contents of the package were explained again in detail by phone or Teams (version 1.6.00.11166; Microsoft Corporation).
3. Third, the participants filled out the cultural probes for 2 weeks, during which the researcher made 2 short phone calls to the participants, being present for possible questions and collecting specific day-based information about the participants' routines and activities related to health and well-being.
4. At the end of the study, the participants were interviewed remotely.
5. Then, they returned the probes package to the researcher in a prepaid envelope. The data collection took a maximum of 5 weeks. When the researcher received the probes package, the participants were sent a small thank-you gift.
6. Data were analyzed.

Cultural Probes

Overview

In this study, cultural probes made data collection of older adults' everyday lives possible in research environments, where remote data collection was important for safety reasons [52]. It helped in collecting data indirectly, as the participants themselves completed the cultural probes at their own pace as part of their everyday lives [42,52,53]. It also helped the participants prepare for the interview [52].

The cultural probe package, recruitment questionnaire, and recruitment letter were tested with 2 nonparticipating members of the target group. The probes, questionnaires, and letter were revised based on the feedback. The research materials were then tested again with another target group member. Finally, the entire study setup with semistructured interviews was tested with 2 target group members.

Background Questionnaires

Among other tasks, the cultural probe package included 4 questionnaires: a background information questionnaire, the Health Confidence Score (HCS) [71], the eHealth Literacy Scale (eHEALS) [72], and the European Health Literacy Survey Questionnaire (HLS-EU-Q16) translated into Finnish by Eronen et al [73]. The other questionnaires were translated into Finnish by the authors. Questions 3 to 10 in the eHEALS questionnaire were used in this study with the added option of "I don't know" [72].

In addition to demographic questions, the background information questionnaire asked about the participant's internet and device use, the eHealth services they had used, and how useful they found them. The purpose of the HCS, eHEALS, and HLS-EU-Q16 questionnaires was to obtain an overview of the participants' capabilities (or perceived skills) to understand health information and use eHealth services.

Diary-Based and Sentence Completion Tasks

The cultural probe package (1) motivated participants to notice and document their actions related to health and well-being as a part of their everyday lives and (2) worked as a triggering element for remote, semistructured interviews. In this remotely conducted study, cultural probes and their exercises helped make the researcher appear more relatable to the participant. This was especially important given the sensitive and personal topics, such as everyday life around health and well-being, discussed in the study, both from the researcher's perspective, and more importantly, from the participant's perspective [74]. The cultural probe tasks included sentence completion tasks [59] and a diary-based task [64-66]. In the diary-based task, participants were asked to make notes regarding their health routines, decisions, and actions for 5 days. In addition, 13 incomplete sentences were formulated following the best practices explained by Nurkka et al [42] and included in the

pretesting phase. In the sentence completion task, participants were asked to complete sentences (Multimedia Appendices 1 and 2) with 4 themes (Table 1).

Table 1. Themes of the sentence completion tasks.

Theme	Included in group 1	Included in group 2
eHealth services now	Yes	Yes
eHealth services in the future	Yes	Yes
Concerns and fears for eHealth services	No	Yes
Informal caregiver's eHealth services (excluded from the data)	Yes	Yes

The cultural probe packages had differences regarding incomplete sentence themes between the group 1 and group 2 participants. The study started with the younger participant group. As data began to accumulate, additional questions addressing concerns and fears were incorporated into the sentence completion tasks, as shown in Table 1. This could support the participants in preparing for the interview as well [52]. However, the interview covered the theme with both participant groups.

Remote, Semistructured Interviews

The themes of the cultural probes were investigated in more detail in remote, semistructured interviews [52], which took 1.5 hours maximum to complete with group 1 participants and 1 hour maximum with group 2 participants. If an interview took longer, the participants were offered the opportunity to continue the interview the next day. The researcher made phone calls using either the Teams call function or a mobile phone, making it easier for participants to take part. Two researchers conducted the interviews, but only 1 conducted each interview for trust-building purposes. The interviews were audio recorded and documented with field notes.

The interviews comprised different perspectives regarding the participants' everyday lives and activities concerning health and well-being. The interview structure (Multimedia Appendix 3) was created based on the themes of the cultural probes. The participants were asked to describe the contents of their cultural probes. The idea was to empower the participants and let them decide how much or in how much detail they were willing to express their thoughts on their well-being and everyday lives. In the interview design, the interview checklist by Tong et al [75] was followed.

Analysis

Overview

After the data were transcribed and pseudonymized, the background questionnaire was analyzed using descriptive statistics, and the qualitative data were analyzed following the inductive content analysis process [69,76,77]. User needs were formulated based on the content analysis of the interviews and the analysis of the cultural probe data. The cultural probe data included the participants' diary notes of their daily well-being activities and completed sentences. The data were analyzed according to data type. The different analysis phases are explained in the subsequent sections.

Preliminary Material Walk-Through

During the interviews, a researcher (PV) identified the preliminary content-based categories and, after each interview, categorized each participant's direct quotes in Excel (version 2208 Build 16.0.15601.20644; Microsoft Corporation). Due to the richness of the data, an additional preliminary review was necessary. The researcher (PV) manually reviewed the entire dataset through after receiving the cultural probes package, using traditional pen-and-paper data analysis methods [78].

After this preliminary data analysis phase, the analysis needs for the next phases were recognized. The analysis needs were identified, keeping in mind that the purpose was to understand user needs. Similar to the data analysis process proposed by Nielsen et al [79], the preliminary material walk-through process needed both systematic and circular practices.

Content Analysis for Interview Data

Afterward, 2 researchers (PV and KS) read the pseudonymized interview transcripts and coded the data in ATLAS.ti 9 (version 9.1.5.0; Scientific Software Development GmbH; Multimedia Appendix 4) to understand the data as a whole and to support the analysis of the cultural probes data later. The process proposed by Mayring [76] was followed in the analysis. When approximately 30% (11/40) of the transcripts were coded, the researchers (PV and KS) revised and compared their categories and agreed on common ones. Finally, the results were discussed again by 2 researchers (PV and KS), and the findings were reported.

Analysis for Background Information Questionnaire Data

The background information questionnaires were analyzed to understand the participants' characteristics and their internet and IT use habits. These questionnaires provided preliminary information about the sample. The questionnaire analysis was done in collaboration between 2 researchers (PV and SK). The means were calculated for the data from each of the 3 questionnaires (PV), and the results were discussed (PV and SK). Owing to the qualitative nature of the study and the small sample size, detailed quantitative analysis was not performed.

Content Analysis for Cultural Probe Data

The data from the cultural probe package were analyzed in collaboration between 3 researchers. Each sentence's main content was analyzed by a researcher (PV), and the results were discussed with 2 other researchers (KS and SK) to ensure the quality of the analysis. The diary analysis focused on the

participants' descriptions of their well-being activities. Activities were written on sticky notes, 1 per note, with the participant code and time of day. After 2 researchers (PV and KS) created the notes for 1 participant, their consistency was checked: a researcher (KS or PV) read the findings and another one (KS or PV) wrote the findings down. The analysis included 644 grouped sticky notes. The sticky notes were grouped by 2 researchers. The notes were taken during the analysis process.

Results

Overview

The results are discussed in three parts: (1) participant information; (2) description of older adults' everyday lives; and (3) user needs, which were expressed based on the findings from the sentence completion and remote, semistructured interview data.

The data comprised 19 cultural probe packages and 329 pages of transcripts of 20 interviews (in Verdana font, 11-point font size, single spaced). One participant did not return the cultural probe package.

Participants

The study included 20 participants (Table 2). The participants were divided into two groups: group 1, those who were still employed, and group 2, those who were retired. Both groups included active computer users with many different devices, but compared to group 1, group 2 included more participants who did not routinely use the internet and IT. However, the variation in device types was greater in group 2 compared to group 1. The results suggest that perhaps retired people use health and well-being services less frequently in general than those who are working.

Table 2. Participant background information (N=20).

	Group 1 (n=12)	Group 2 (n=8)
Age (y)	55 to 74	>75
Gender, n (%)		
Women	9 (75)	6 (75)
Men	3 (25)	2 (25)
Internet activity, n (%)		
Internet use daily or many times in a day	11 (92)	6 (75)
Internet use on a weekly basis	1 (8)	1 (12)
Do not know to use or do not use the internet	0 (0)	1 (12)
Internet habits, n (%)		
Most popular devices for internet use		
Computer	12 (100)	5 (63)
Cell phone	12 (100)	4 (50)
Tablet	9 (75)	4 (50)
Place of internet use		
At home	12 (100)	7 (88)
Outside home	10 (83)	3 (38)
The most popular reasons to use the internet		
Searching information	12 (100)	7 (88)
Banking	12 (100)	6 (75)
Health and well-being services	12 (100)	4 (50)
eHealth use, n (%)		
Most popular digital public health care services used		
Patient portal: checking health information and receipt renewal	12 (100)	4 (50)
Appointment booking	11 (92)	4 (50)
Chat, email, or SMS text message to health care professionals	8 (67)	—[a]
Visiting a health center's web pages	—	2 (25)

[a]Data not available.

All group 1 participants (12/12, 100%) and half (4/8, 50%) of the group 2 participants checked their health information and renewed their prescriptions via eHealth services. Internet users used it for banking and information searches, and 80% (16/20) of them also used it for health services.

Multimedia Appendix 5 presents the background questionnaire results. The eHEALS results (20/20, 100%) show the participants' familiarity with using digitally offered health information, with differences between the participant groups. The results of the HLS-EU-Q16 (19/20, 95%) on health literacy indicate that the participants seemed to have at least basic health literacy skills. One participant did not send the questionnaire back to the researchers. The HCS results (20/20, 100%) on patients' confidence in taking care of their own health show that the participants had at least a basic understanding of their health situation and how to operate in case of health challenges [63].

Everyday Life of Older Adults

The analysis of the data from the diary-based task and the interviews resulted in 15 themes in the participants' everyday activities. The themes were spending time at home, dining and cooking, routines, hobbies or exercising, sleep, mundane activities, work-related activities, medical treatment, cottage or nature, friends, pets, pampering and rest, communication with a relative, planning and controlling everyday life, and weather.

The most common activity was spending time at home, often watching television (102/622, 16.4%); spending time at home and dining and cooking (96/622, 15.4%) were the most popular regular activities of the participants. The importance of food for well-being was often mentioned in the data. Exercising was also popular (72/622, 11.6%), with walking and jogging being commonly practiced, including Nordic walking and climbing stairs. Walking could also be a social activity with a partner or a dedicated group. Sleep was mentioned in the data 51 out of 622 times (8.2%), as were mundane activities, such as going to the post office or having lunch or a coffee (51/622, 8.2%). The repetitive nature of routines, including spending time at home, the importance of food, and the variety of exercises, is reflected in the following quotes (the quotes are translated from Finnish):

> *From Monday to Friday, I wake up 5.50 am on two mornings, and on other mornings 6.50 am. And that 5.50 am means, that I go swimming. And on other mornings I eat my breakfast normally at home and go to work. I drink my morning coffee at the office.* [Group 1, participant U2]

> *There have been such beautiful days lately. Today it seems it is not so beautiful, or at least the sun is not shining yet. I feel I'm privileged that I'm able to live in my hometown, because this is such a beautiful city. I started with 30 min walking exercises, and now I do my walking exercise two times in week. Each walking exercise takes one and half hours nowadays. I explore those different terrains, and we have here a lot of a lot of stairs and a lot of small hills, which I use to improve my fitness.* [Group 1, participant U12]

> *I start my work at 9 am, then I eat breakfast, and walk with my dogs. I normally work at my computer until 6–8 pm. I eat, and drink coffee during short work breaks. At 4 pm I give food to my dogs. In the evening, I don't meet anybody else than other dog owners. I also meet new employees during my workday.* [Group 1, participant U6]

> *I eat breakfast in the morning, and at the same time, I read news from my iPad. My husband and I have not ordered any paper newspapers.* [Group 2, participant HX4]

> *I might visit quite often to meet my neighbor, but not on an everyday basis. But it is a part of my everyday life. ...And then, I go quite often after 6 pm a short walk, and at the same time, visit at my friend, who has two cats. ...Then I make some dinner and watch TV News. I do not usually follow any TV series, but I might sometimes watch something from TV. I go to sleep at around 12 am.* [Group 2, participant HX5]

> *I don't have so many meetings anymore, but my week starts with English lessons every Monday at 10 am. At 4 pm I go to an hour's outdoor exercise. In the evening, I often go to German language speaking exercise group. Next day, I go to aqua jogging, and after that we take a cup of coffee with the members of that group. Next day I participate in outdoor walking exercise group, with which we normally walk an hour together. And at the end of the week, I often do some housework, a little bit cleaning home or something.* [Group 2, participant U7]

Watching television was emphasized in both groups, and activities related to cleaning and housekeeping were popular. Overall, participants in both groups seemed to like routines. The everyday life activity themes are detailed in Multimedia Appendix 6.

The Needs of Older Adults for eHealth Services

Overview

The older adults' needs for eHealth services were collected based on the main findings from remote, semistructured interviews and sentence completion tasks. The themes discovered from the interviews, the main findings under each theme, and the participants' needs based on those findings are provided in Multimedia Appendix 7. In addition, sentence completion revealed the participants' wishes and needs for eHealth services in the future (Multimedia Appendices 1 and 2). The needs described in the following sections were expressed based on the data.

The Need for a Carefree Mind

Older adults need a carefree mind when taking care of their health and regarding their current health and well-being situation, which can be supported with eHealth services. They need to understand the safety of eHealth service use and be able to manage health-related matters well regardless of the channel (eg, mobile phone, chat, or remote appointment). Users' technical skills vary, and some use technical devices versatilely. Support and training possibilities help with this need. Feedback

on one's current health and well-being situation via eHealth services is appreciated. The safety needs are evident in the following quotes:

> *I avoid using all parts [of the user interface] that have even a little bit of a foreign language to me. Because it might happen there that I don't understand something and make a wrong choice somehow [in the user interface].* [Group 1, participant U2]

> *I think managing my health-related affairs should be safe.* [Group 1, participant U6]

The need to obtain health care service regardless of the service channel is reflected, for example, in the following comments:

> *I would like to be able to use eHealth services in parallel with other service channels in the future.* [Group 1, participant U1]

> *I look for information on the internet every day. I use all possible medical and health services that are available digitally. And the same thing with public services: everything that is possible, I do them in digital channels.* [Group 2, participant H3]

On the basis of the data, people's diversity is emphasized: being able to use the eHealth service can be a big challenge for one person, but for another, it brings benefits and ease. Owing to older adults' varying technology skills and device availability, health self-management was desired in varying ways and different channels, either face-to-face or digitally. In addition, group 1 participants seemed to be experienced or very experienced with technology and felt very comfortable using eHealth services. The clear variations between the participant groups should be considered in all design decisions regarding eHealth service development.

Varying skills and comfort levels with using eHealth services are demonstrated, for example, in the following quotes:

> *For me, using eHealth services is natural and easy.* [Group 1, participant U6]

> *For me, using eHealth services is easy and self-evident.* [Group 2, participant HX4]

> *eHealth services do not help me because I have neither the equipment nor the skills.* [Group 2, participant S8]

> *Learning something new [about digital services] is difficult because of the tricky terminology. I've never studied English at school, and systems use a special system language. ...Then these jungles of safety encryptions and passwords: remembering them is almost the most difficult part of it.* [Group 2, participant S9]

The Need for eHealth Service Adaptation Based on Current Life Situation, Use Contexts, and Everyday Life

Regarding different life situations, the participants needed support via eHealth with their health and well-being in changing situations, such as when retiring, regardless of their current technology skills or devices. eHealth could help them, for example, with mental well-being or keeping physical activity levels high enough. The changing situations are evident in the following quotes:

> *My working years are in the final phase. It's sure that I have a couple of years left until I retire officially. After that, my children are already adults as well. But currently, my life is quite work-oriented, long workdays, and I'm all the time busy.* [Group 1, participant U1]

> *I have done more or less work [in my life]. And now I'm retiring. That's a pretty big change.* [Group 1, participant U4]

> *I'd like to use eHealth services in the future even after I retire.* [Group 1, participant U12]

> *I have a little bit tricky stage of life. My son went to heaven two weeks ago, and I am currently going through grief. Five years ago, my husband died. So, that kind of stage of life I have now.* [Group 2, participant H2]

> *Yes, it was easier when I was healthy. After all, then I could do what I wanted and run if I liked and, in every way. ...That life was completely different then. I can't say, however, that none of these moments in this life have been completely unpleasant. ...When I think about this end of life, yes, I have had a good life the whole time, no matter what era it was. They have all been part of life that was then, and it was good at that moment.* [Group 2, participant S7]

Older adults desired eHealth services that could adapt to their current and evolving life situations, such as retirement or changes in health. These services should be tailored to their lives in terms of both content and functionality, including the user interface, and should be flexible enough to adjust according to the season, regardless of their location. Older adults want to take care of their health and well-being regardless of location (from places other than home or a clinic, such as a summer cottage). Service adaptation to everyday life makes eHealth services relevant.

However, the ease of use of an eHealth service becomes relevant only after it has been ensured that it can be accessed at all. One's health state affects everyday life and the ability and willingness to use eHealth services. For example, when a health situation improves due to following healthy daily routines, the content of the eHealth service should follow this new health situation and its user's new needs for the eHealth service. eHealth services should be adaptable to changes in life and health situations as well as fluctuations in motivation to use eHealth, both in the short and long term. eHealth service's role and value as a part of changing life situations and user needs regarding eHealth service's adaptability are reflected in the following quotes:

> *eHealth services help me very well in my health situation.* [Group 1, participant U12]

> *eHealth services help me plan schedules flexibly.* [Group 1, participant U7]

> *For me, the most important thing with eHealth services is that I don't have to commit to a specific time.* [Group 1, participant U10]

For me, the most important thing in eHealth services is the right information at the right time. [Group 2, participant S9]

For me, the most important thing with eHealth services is that I get information immediately without waiting. [Group 2, participant HX4]

Using them [eHealth services] saves time. [Group 1, participant U8]

The Need for a Holistic Perspective

Older adults need content that supports their health and well-being holistically. eHealth services should, for example, include perspectives on their current everyday lives and health situations. Group 1 members often wished that eHealth services would be one service among others, smoothly integrated into everyday life. For group 2 members, more research is needed to investigate their perspectives on a suitable package of service channels and content so that they receive the same holistic support as group 1 members. The example quotes behind the need for holistically offered health and well-being support as a part of older adults' everyday lives are as follows:

I fill in crosswords, and I read when I like to. I cook, but not every day now. Things like that I get pleasure from. After all, all in all, the fact that there are not so many financial worries is a part of that balance. Everything you need is here and now. Everyone can ask themselves the question whether they are happy. [Group 1, participant U11]

For me, the most important thing in eHealth services are the right health goals considering the right information; based on researched and measured information. [Group 1, participant U7]

eHealth services help me to plan flexible schedules. [Group 1, participant U7]

eHealth services help me take care of things related to my health at once. They save time. I get e.g., lab results through a digital service. [Group 1, participant U13]

eHealth services help me get an overall picture of my health. [Group 1, participant U1]

I read a lot. Now that I've been here sick and the moments, I've been awake, I'm reading something all the time. [Group 2, participant HX6]

And then jogging with the dog. It keeps me in balance. All my social interaction now takes place there, walking the dog, because there are many dogs here, and we have lived here 45 years, so there are so many people who know dogs. [Group 2, participant S8]

The Need to be Able to Avoid Using eHealth Services But to Use Specific eHealth Functions

Older adults, especially group 2 members, need to be able to select the service channel: eHealth services or face-to-face. On the other hand, there was a need for eHealth functions. For example, group 2 members often appreciated the opportunity to look at laboratory results without having to go anywhere. The COVID-19 pandemic era might have affected this result. The following example quotes bring insights into the older adults' needs for specific functions and the possibility of obtaining health care services via service channels other than eHealth services:

In principle, they [service providers] accept to do business [face-to-face], but if you ask for some advice there, they recommend that you "go to do it online or look it up online." They don't understand that not everything works perfectly on the internet for them either. I think it's pointless for them to guide you to do business online when it can't be done there online. [Group 1, participant U8]

Or you can use a video connection to contact the nurse! They are certainly good additions to the healthcare service. [Group 1, participant U1]

It's embarrassing that I don't know how to book appointments [digitally]. I don't know how to book a time for our laboratory or X-ray. I don't know how to make an appointment. And you should know that, but when there was no need, you didn't learn. [Group 2, participant HX6]

When I imagine myself retired, and when I no longer have a work computer, and if it is difficult to get from one place to another, then of course, they [eHealth services] will be useful. And then the chat service, where you can easily ask for help or advice!

I think managing my health-related affairs should be possible in person. [Group 2, participant S8]

This means that eHealth services should be accessible: (1) devices should be available and (2) the network connection should be stable enough, and the service must be available regardless of time. The service should be offered in a multichannel manner, including personal contact, phone service, or in writing (eg, chat).

Operations related to health and well-being, such as appointment booking through eHealth services, must be clear and easy to use for all user groups. The operations and the eHealth service should support long-term and short-term health and well-being planning in terms of the content and functions available to both end users and health care professionals. It should work reliably in all situations and support older adults' empowerment in managing and controlling their current and future health situations. This is evident, for example, in the following quotes:

I think operations related to my health and well-being should be easy. [Group 2, participant HX5]

eHealth services do not help me at all. [Group 2, participant HX6]

User Needs Regarding User Experiences

The completed sentences revealed that the older adults' experiences with eHealth services varied, with some finding them easy to use and others finding them challenging. The use of eHealth services was felt to be part of everyday life and routines, and it did not produce many "peak experiences," which are the most memorable experiences producing positive

emotions [70,80]. The variation in user experiences from eHealth use is apparent in the following example quotes:

> *For me, using eHealth services is easy, and a routine.* [Group 1, participant U13]
>
> *Using them [eHealth services] makes me a little irritated.* [Group 1, participant U1]
>
> *I think managing my health-related affairs should be confidential, and the response quick and empathetic.* [Group 1, participant U11]
>
> *Using them [eHealth services] puts me in a bad mood.* [Group 2, participant HX6]

Two (25%) out of 8 participants in group 2 had worrying experiences, such as a loss of self-confidence in the use of eHealth services and an increase in concern about the use of eHealth services. These worries are reflected in the following example quotes:

> *Using them [eHealth services] makes me feel insecure.* [Group 2, participant S7]
>
> *The thing that worries or scares me the most about eHealth services is that soon you won't be able to see a doctor anymore when everything will be done digitally.* [Group 2, participant S8]

However, eHealth services especially brought peace of mind to group 1 participants, seeming to improve their well-being. For both groups, eHealth services produced important information about their current health situation. eHealth services seemed to be used in both age groups, not only out of necessity but also to obtain information and improve the flexibility of taking care of their health. eHealth services provided longer-term insights into their own health status, and information obtained through eHealth services was highlighted at different points. Finding out about their own health status seemed important. With the help of eHealth services, the level of concern about their own health condition improved. Group 2 members had more negative views of eHealth services, and they also mentioned a lack of equipment as a barrier to eHealth service use. eHealth service's role in offering information, flexibility in taking care of one's health, or role as a calming factor is reflected in the following example quotes:

> *I use eHealth services if I book an appointment with my doctor, or check my health information* [Group 1, participant U12]
>
> *I use eHealth services whenever possible.* [Group 1, participant U6]
>
> *Using them gives me peace of mind when I can book appointments and find out about health-related issues at a time that suits me.* [Group 1, participant U13]

Other barriers to eHealth services were lack of human contact, inefficiency, and challenges with its use. eHealth services divided participant opinions regarding its benefits: group 1 participants did not have as many helpful experiences with eHealth services as group 2 participants. Both groups included individuals who felt that eHealth services did not provide holistic support during life changes. At the same time, group 2 participants felt that eHealth helped them acquire information about health-related issues, supported them in planning health-related issues, and helped them commit to lifestyle changes. On the other hand, they lacked competence and felt that eHealth services (use processes and user interfaces) was too complicated. The negative experiences and use barriers are evident in the following quotes:

> *With eHealth services, I am saddened by the lack of consistency and aimlessness, as well as the lack of genuine and cheerful support.* [Group 1, participant U7]
>
> *With eHealth services, I am frustrated by delays and waiting.* [Group 1, participant U6]
>
> *With eHealth services, I am saddened by my laziness and lack of time available to study the use of eHealth services.* [Group 2, participant H3]
>
> *With eHealth services, it saddens me that I don't know how to use them.* [Group 2, participant HX6]

The most important things in eHealth services use were efficiency, ease and smoothness of transactions, access to information, planning (where group 1 members had diverse answers), and support for well-being. Participants in group 2 had the same eHealth needs as those in group 1 but they were more skeptical (Multimedia Appendix 8).

The efficiency, ease, and support for well-being via eHealth services are apparent in the following quotes:

> *Using them [eHealth services] saves my time.* [Group 1, participant U6]
>
> *Using them [eHealth services] makes me think about my health based on researched information.* [Group 1, participant U10]
>
> *For me, the most important thing about eHealth services is that I don't have to commit to a specific time or date.* [Group 1, participant U10]
>
> *eHealth services help me renewing my medicine prescriptions.* [Group 2, participant S7]

Discussion

This qualitative exploratory study aimed to gain a more in-depth understanding of older adults' needs for eHealth services as part of their everyday lives in different life stages. The data were collected with interviews and a cultural probe package, including 4 background information questionnaires, sentence completion tasks, and a diary-based exercise.

Principal Findings

Everyday Life of Older Adults

Both participant groups valued time spent at home and home-based activities as part of their everyday lives. Healthy eating habits were an especially important aspect of the participants' well-being from their own perspective. These older adults enjoyed different activities, such as jogging and Nordic walking, which is in line with other literature [31].

Older adults generally appreciated bringing health and well-being into everyday life. They had weekly or daily

activities or hobbies that supported them to reach their long-term health targets. These routines varied greatly, including exercise, good night sleep, personal hygiene, repetitive food, pet care, and housework-related morning routines. Group 1 participants reported routines and activities that partly differed from those of group 2 participants. For example, all the routines related to the gym or exercising while commuting were reported by group 1 participants. Interestingly, the only mention of reading newspapers with a tablet was reported by a participant from group 2. Furthermore, a mention of doing morning exercise with television was reported by a group 2 participant. Cleaning routines were mainly mentioned by participants from group 2.

The eHealth service can both support health self-management on a long-term basis and activate and support older adults in ad hoc type of activities and operations related to health and well-being, for example, by offering information based on the current health situation or reminding them of a (daily or weekly) hobby.

The Needs of Older Adults for eHealth Services

On the basis of the interviews and sentence completion tasks, four needs for eHealth services were identified: (1) the need for a carefree mind; (2) the need for eHealth service adaptation based on the current life situation, use contexts, and everyday life; (3) the need for a holistic perspective; and (4) the need to avoid using eHealth services but to use specific eHealth service functions. These needs should be considered when designing eHealth services that support older adults' healthy living. In addition, they generally apply to both user groups but may be emphasized differently depending on the group.

On the basis of our findings, the older adults seemed to have generally positive attitudes toward eHealth, aligning with those identified by Cabrita et al [15] and Mielonen et al [28]. The infrastructure behind eHealth services must be stable; then, eHealth services can be offered in an accessible way and without limitations for everyone, assuming that end devices are available. Owing to varying end devices among older adults, eHealth services should be scalable regardless of the service channel or devices. Therefore, the design should support the reliability of eHealth services so that they work in all situations, supporting older adults' health self-management [15,70]. Our study adds the perspective of changing time spans: the service must function smoothly for both short-term and long-term use, accommodating variations in use duration.

Well-being, sickness, health, and illness seem to be subjective personal experiences. The need for eHealth services changes due to the user's current health and well-being situation: eHealth services become unnecessary when health and well-being increase. Therefore, it must be possible for older adults to easily try out whether eHealth services will benefit them. An eHealth service may also be used less frequently at times.

eHealth services can support older adults' health and well-being, particularly when the content of the service aligns with their everyday lives and interests. Therefore, we recommend that the eHealth service work as part of an older adult's healthy routines, which might include, for example, encouraging attitudes and functions for housework and other activities.

Comparison With Prior Work

Our results show that understanding the user's life stage and the key interests at that stage is important, as found in other literature [47]. Owing to older adults' changing health and life situations [1-3,10,11,47], an individual eHealth service should function as part of a wider service offering so that the service supports and adapts to changing life situations and health statuses. For example, in this study, it was recognized that both those who are still working and those who are retired use IT and eHealth services. IT use at work might support learning new skills and sometimes even updating them [70]. In addition, eHealth service flexibility is valued [4]. The need for eHealth services to adapt to the older adult's current life and health situation is in line with the study by Reiners et al [81].

Earlier research recognized a large variation in activities that support older adults' well-being, in line with our findings: exercising; cooking; and hobbies, such as language courses [82]. Even with poor health, older adults achieve well-being and positive effects from daily activities [82,83]. In addition to physical activities, daily activities, such as reading books or watching television, can also offer relaxation [81]. Owing to the differing routines and daily activities, the need for eHealth adaptation based on the current life situation, use contexts, and everyday life was recognized in this study, and we recommend that eHealth services work as a part of older adults' daily living smoothly. Previous research supports our results behind this user need: older adults are a heterogeneous group whose activity possibilities vary regarding their living conditions or current health situation [31]. However, the importance of studying the use of eHealth service as part of everyday life has been recognized [19].

The devices with which the eHealth services are used should be considered when designing these services. Given the possible effect of the small sample size, it can be seen from the data that group 2 participants seemed to use eHealth services somewhat less than group 1 participants. One reason for that could be that older adults who are still employed might use eHealth services for work reasons, as well [70]. On the other hand, it can probably be expected that, nowadays, older adults who are retired have used at least some IT at work.

Understanding the safety of eHealth services use was important to older adults, as in earlier research [4,5,10,18]. In line with previous literature [18,70], they needed available support and training for eHealth service use. In particular, the older participants in this study also needed not to be forced to use the eHealth service. However, this varied considerably between participants, which supports earlier research results [18].

In line with an earlier study [15], one of the most important design functions is to ensure that the eHealth service conveys empathy to older adults. This can be offered via an empathic tone of voice in the services and by offering them through service channels that are especially suitable to each older adult's current situation. The whole life span, including transitions to different life phases, such as retirement, should be considered in design decisions [2,47,70]. This can be done, for example, by offering relevant content related to health and well-being to

different older adult groups and by making sure that the content is relevant to each end user's unique life and health situation.

On the basis of our findings, both those still working and those who have already retired have suitable devices and sufficient technical skills, as also reported by Mielonen et al [28]. However, eHealth services were used only by half (4/8, 50%) of the older participant group. Our findings support the observations of Kruse et al [3] regarding common barriers to eHealth use among older adults: a lack of self-confidence, a fear of making irreversible mistakes, and content that does not meet their needs. Fear can be dispelled with a clearer service concept, implementation support, user support, and instructions, as well as reassurance, such as by highlighting safety both in the service content and in the presentation of the service. In addition to minimizing fear, strengthening the self-confidence of the end user of the service plays an important role. Another phenomenon that emerged from the data was that well-being, health, and illness seem to be subjective, personal experiences.

The need for eHealth services changes due to the user's current health and well-being situation; eHealth services become unnecessary when health and well-being improve. Therefore, it should be possible for older adults to easily try out whether the eHealth service will benefit them. This can be done, for example, by offering trial versions of services, which other research [15,18,84] recommends. Being able to safely try out an eHealth service could help reduce fears, build self-confidence, and better identify the current need for the service. Another helpful option would be to group service functions into easier basic functions and possibly other more complex functions that are recommended to be familiarized with time, as is recognized in the literature [15,47]. However, based on the study findings, the end user's age or life situation does not appear to be a decisive factor in determining the need for user interface content. That said, older adults who are retired may face slightly more challenges in using eHealth services compared to those who are still working.

As also found in other literature [18,70], individual eHealth service should function as part of a wider, multichannel service package so that it adapts to changing life situations and changes in health status. It should work as a channel for end users to receive empathy in challenging health situations. This could be done with effective but sensitive communication possibilities and with empathic content. Because eHealth services should support older adults' health and well-being, it is important to look at and connect their well-being in the IT context [84]. This can be done according to content and with easy and logically functioning user interfaces. In addition, this study's participants wished for an option to not use eHealth services or specific eHealth functions. This need is in line with other research findings [85].

Limitations

This qualitative exploratory study aimed to gain a more in-depth understanding of older adults' needs. The sample was small; the participants were from 1 country, and they were likely to be more interested than average in eHealth services. Older adults are a diverse group from the perspective of, for example, cognitive, motor, or technology use abilities [61]. Therefore, completely covering the population with a representative sample was not possible. This has potentially biased the results. In addition, the COVID-19 pandemic era could have influenced the results, but this was not especially followed as a part of this study.

Regarding the adaptation of eHealth services to different life and health situations, we do not know whether the service should always be the same or differ depending, for example, on whether the health situation has changed, in which case the role and meaning of the service changes from supporting well-being to treating illness. We recommend this as a future research area, especially in long-term studies. More research is needed on the best service channel combinations for older adults to obtain sufficient holistic health and well-being support. In the future, it would be worthwhile to test the design recommendations with representatives of the designer community and older adults, for example, in co-design workshops. In this study, the older adults' needs in different life situations were not compared to those of other age groups. However, this would be valuable to study in the future.

From a methodological perspective, the cultural probe package, including diary-based and sentence completion tasks, combined with remotely conducted semistructured interviews seemed to be effective in collecting information about participants' everyday lives and in mapping user needs for eHealth services. This is interesting because, as far as we know, not many cultural probes studies on this topic are available. The working participants completed more handwriting-based sentences. Therefore, cultural probes including many handwriting exercises might fit better for them. On the other hand, the data were collected using several methods, which confirmed the findings of the study and helped identify user needs. The methods seemed to work well in remote conditions during the COVID-19 pandemic era.

Conclusions

In this qualitative study, we collected information about the everyday lives of older adults in different life situations and identified their related needs for eHealth services. The results helped us understand how older adults see and experience health care with the help of eHealth services as part of everyday life. The older adults in this study often preferred a consistent routine in their daily practices including healthy eating and exercise. eHealth services could offer time saving and flexibility within these routines. A lack of devices or skills was often mentioned as a use barrier. On the basis of the results, four main needs of the older adults for eHealth services were identified: (1) the need for a carefree mind, in which older adults could receive status information on their current health and well-being situation via eHealth services; (2) the need for eHealth service adaptation based on the current life situation, use contexts, and everyday life; (3) the need for a holistic perspective, in which older adults wished to receive support for health issues from many perspectives and through several different service channels; and (4) the need to avoid using eHealth services but to use specific eHealth service functions, in which face-to-face support for their health was especially appreciated. On the basis of the results, eHealth services should be designed such that

they fit well in older adults' everyday lives and adapt to users' everyday practices and health statuses according to content. eHealth services should be accessible and reliable.

Acknowledgments

The authors would especially like to thank the study participants. The authors would also like to thank Marina Broman for completing 3 of the interviews and Marjaana Kanervo, Annetta Kurvinen, Tytti Natunen, Kirsi Pitkälä, Sanna Pussinen, Semra Räikkönen, Reija Saarinen, Sanna Solala, Arja Sunila, and Iina Vartia for the help with the research arrangements and collaboration. The work was supported by the Strategic Research Council at the Academy of Finland (grants 352501 and 352503) and NordForsk (project 100477).

Data Availability

The datasets generated and analyzed during this study are not publicly available due to the sensitive nature of the data, but the numeric data are available from the corresponding author on reasonable request.

Authors' Contributions

The study was conceptualized by PV, SK, and RRH. PV was involved in the methodology, investigation, project administration, data curation, visualization, and formal analysis of the study. KS contributed to the formal analysis and validation of the study. PV, KS, and SK were involved in preparing the original draft of the study. SK contributed to the methodology, resources, data analysis, validation, supervision, and funding acquisition of the study. RRH also contributed to the resource collection of the study. All authors were involved in writing, reviewing, and editing the manuscript.

The visualizations used in the cultural probes were under creative commons licenses. The creators' nick names were OpenClipart-Vectors, Emmie_Norfolk, Buntysmum, Clker-Free-Vector-Images, bibin9363pbr, 1,820,796, ArtsyBee, and b0red.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sentence completion examples.
[PDF File (Adobe PDF File), 191 KB - humanfactors_v12i1e50329_app1.pdf]

Multimedia Appendix 2
Sentence completion task by age group (current situation).
[PDF File (Adobe PDF File), 138 KB - humanfactors_v12i1e50329_app2.pdf]

Multimedia Appendix 3
Remote, semistructured interview (translated from Finnish).
[PDF File (Adobe PDF File), 121 KB - humanfactors_v12i1e50329_app3.pdf]

Multimedia Appendix 4
Interview data analysis code comparison and final codes after comparison.
[PDF File (Adobe PDF File), 166 KB - humanfactors_v12i1e50329_app4.pdf]

Multimedia Appendix 5
Health Confidence Score, eHealth Literacy Scale, and European Health Literacy Survey Questionnaire results by age group.
[PDF File (Adobe PDF File), 128 KB - humanfactors_v12i1e50329_app5.pdf]

Multimedia Appendix 6
Identified themes in life activities.
[PDF File (Adobe PDF File), 80 KB - humanfactors_v12i1e50329_app6.pdf]

Multimedia Appendix 7
Interview themes, example quotes, and user needs.
[PDF File (Adobe PDF File), 208 KB - humanfactors_v12i1e50329_app7.pdf]

Multimedia Appendix 8
Differences in the completed sentences between the 2 groups, focusing on the situation in the future.
[PDF File (Adobe PDF File), 131 KB - humanfactors_v12i1e50329_app8.pdf]

References

1. Czaja SJ, Lee CC. The impact of aging on access to technology. Univ Access Inf Soc 2006 Dec 8;5(4):341-349. [doi: 10.1007/s10209-006-0060-x]
2. Kornadt AE, Kessler EM, Wurm S, Bowen CE, Gabrian M, Klusmann V. Views on ageing: a lifespan perspective. Eur J Ageing 2020 Dec 11;17(4):387-401 [FREE Full text] [doi: 10.1007/s10433-019-00535-9] [Medline: 33380996]
3. Kruse C, Fohn J, Wilson N, Nunez Patlan E, Zipp S, Mileski M. Utilization barriers and medical outcomes commensurate with the use of telehealth among older adults: systematic review. JMIR Med Inform 2020 Aug 12;8(8):e20359 [FREE Full text] [doi: 10.2196/20359] [Medline: 32784177]
4. Hechinger M, Hentschel D, Aumer C, Rester C. A conceptual model of experiences with digital technologies in aging in place: qualitative systematic review and meta-synthesis. JMIR Aging 2022 Sep 09;5(3):e34872 [FREE Full text] [doi: 10.2196/34872] [Medline: 36083625]
5. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
6. Allmér HR. Digital wellness services' servicescape for young elderly. In: Proceedings of the 31th Bled eConference Digital Transformation - Meeting the Challanges. 2018 Presented at: BLED 2018; June 17-23, 2018; Bled, Slovenia URL: https://aisel.aisnet.org/bled2018/32/ [doi: 10.18690/978-961-286-170-4.10]
7. Carlsson C, Walden P. Digital wellness services: key to better quality of life for young elderly authors. In: Proceedings of the 31th Bled eConference Digital Transformation - Meeting the Challanges. 2018 Presented at: BLED 2018; June 17-23, 2018; Bled, Slovenia URL: https://aisel.aisnet.org/bled2018/26/ [doi: 10.18690/978-961-286-170-4.16]
8. Ageing and health. World Health Organization. 2024 Oct 1. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-12-01]
9. MacNee W, Rabinovich RA, Choudhury G. Ageing and the border between health and disease. Eur Respir J 2014 Nov;44(5):1332-1352 [FREE Full text] [doi: 10.1183/09031936.00134014] [Medline: 25323246]
10. Steele Gray C, Miller D, Kuluski K, Cott C. Tying eHealth tools to patient needs: exploring the use of eHealth for community-dwelling patients with complex chronic disease and disability. JMIR Res Protoc 2014 Nov 26;3(4):e67 [FREE Full text] [doi: 10.2196/resprot.3500] [Medline: 25428028]
11. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul 07;380(9836):37-43 [FREE Full text] [doi: 10.1016/S0140-6736(12)60240-2] [Medline: 22579043]
12. Li Y, Schoufour J, Wang DD, Dhana K, Pan A, Liu X, et al. Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: prospective cohort study. BMJ 2020 Jan 08;368:l6669 [FREE Full text] [doi: 10.1136/bmj.l6669] [Medline: 31915124]
13. Mehta N, Myrskylä M. The population health benefits of a healthy lifestyle: life expectancy increased and onset of disability delayed. Health Aff (Millwood) 2017 Jul 19:1495-1502 [FREE Full text] [doi: 10.1377/hlthaff.2016.1569] [Medline: 28724530]
14. Chudasama YV, Khunti K, Gillies CL, Dhalwani NN, Davies MJ, Yates T, et al. Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: a longitudinal cohort study. PLoS Med 2020 Sep;17(9):e1003332 [FREE Full text] [doi: 10.1371/journal.pmed.1003332] [Medline: 32960883]
15. Cabrita M, Tabak M, Vollenbroek-Hutten MM. Older adults' attitudes toward ambulatory technology to support monitoring and coaching of healthy behaviors: qualitative study. JMIR Aging 2019 Mar 12;2(1):e10476 [FREE Full text] [doi: 10.2196/10476] [Medline: 31518252]
16. Heponiemi T, Jormanainen V, Leemann L, Manderbacka K, Aalto AM, Hyppönen H. Digital divide in perceived benefits of online health care and social welfare services: national cross-sectional survey study. J Med Internet Res 2020 Jul 07;22(7):e17616 [FREE Full text] [doi: 10.2196/17616] [Medline: 32673218]
17. Wilson J, Heinsch M, Betts D, Booth D, Kay-Lambkin F. Barriers and facilitators to the use of e-health by older adults: a scoping review. BMC Public Health 2021 Aug 17;21(1):1556 [FREE Full text] [doi: 10.1186/s12889-021-11623-w] [Medline: 34399716]
18. Kaihlanen AM, Virtanen L, Buchert U, Safarov N, Valkonen P, Hietapakka L, et al. Towards digital health equity - a qualitative study of the challenges experienced by vulnerable groups in using digital health services in the COVID-19 era. BMC Health Serv Res 2022 Feb 12;22(1):188 [FREE Full text] [doi: 10.1186/s12913-022-07584-4] [Medline: 35151302]
19. Hirvonen N, Enwald H, Känsäkoski H, Eriksson-Backa K, Nguyen H, Huhta AM, et al. Older adults' views on eHealth services: a systematic review of scientific journal articles. Int J Med Inform 2020 Mar;135:104031. [doi: 10.1016/j.ijmedinf.2019.104031] [Medline: 31918340]

20. Grindrod KA, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]
21. Scheibe M, Reichelt J, Bellmann M, Kirch W. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. Med 2 0 2015 Mar 02;4(1):e1 [FREE Full text] [doi: 10.2196/med20.3912] [Medline: 25733033]
22. Williams V, Price J, Hardinge M, Tarassenko L, Farmer A. Using a mobile health application to support self-management in COPD: a qualitative study. Br J Gen Pract 2014 Jul;64(624):e392-e400 [FREE Full text] [doi: 10.3399/bjgp14X680473] [Medline: 24982491]
23. Seifert A. The digital exclusion of older adults during the COVID-19 pandemic. J Gerontol Soc Work 2020;63(6-7):674-676. [doi: 10.1080/01634372.2020.1764687] [Medline: 32401181]
24. Seifert A, Hofer M, Rössel J. Older adults' perceived sense of social exclusion from the digital world. Educ Gerontol 2019 Feb 08;44(12):775-785. [doi: 10.1080/03601277.2019.1574415]
25. Rice M, Newell A, Morgan M. Forum Theatre as a requirements gathering methodology in the design of a home telecommunication system for older adults. Behav Inf Technol 2007 Jul;26(4):323-331. [doi: 10.1080/01449290601177045]
26. Mannheim I, Schwartz E, Xi W, Buttigieg SC, McDonnell-Naughton M, Wouters EJ, et al. Inclusion of older adults in the research and design of digital technology. Int J Environ Res Public Health 2019 Oct 02;16(19):3718 [FREE Full text] [doi: 10.3390/ijerph16193718] [Medline: 31581632]
27. Mannheim I, Wouters EJ, Köttl H, van Boekel LC, Brankaert R, van Zaalen Y. Ageism in the discourse and practice of designing digital technology for older persons: a scoping review. Gerontologist 2023 Aug 24;63(7):1188-1200 [FREE Full text] [doi: 10.1093/geront/gnac144] [Medline: 36130318]
28. Mielonen J, Kinnunen UM, Saranto K, Kemppi A, Kuusisto H. The effect of chronic diseases on the use of health technology and digital services in the elderly population in Finland. Stud Health Technol Inform 2020 Nov 23;275:147-151. [doi: 10.3233/SHTI200712] [Medline: 33227758]
29. Schweitzer J, Synowiec C. The economics of eHealth. Results for Development Institute. 2010. URL: https://r4d.org/wp-content/uploads/Economics-of-eHealth.final_.3Nov1010_0.pdf [accessed 2024-12-01]
30. Stephanidis CC, Salvendy G, Antona M, Chen JY, Dong J, Duffy VG, et al. Seven HCI grand challenges. Int J Hum Comput Interact 2019 Jul 01;35(14):1229-1269. [doi: 10.1080/10447318.2019.1619259]
31. Eurostat. Ageing Europe: looking at the lives of older people in the EU: 2020 edition. Publications Office of the European Union. 2020. URL: https://ec.europa.eu/eurostat/web/products-statistical-books/-/ks-02-20-655 [accessed 2023-01-13]
32. Valokivi H, Carlo S, Kvist E, Outila M. Digital ageing in Europe: a comparative analysis of Italian, Finnish and Swedish national policies on eHealth. Ageing Soc 2021 Jul 08;43(4):835-856. [doi: 10.1017/s0144686x21000945]
33. Vines J, Pritchard G, Wright P, Olivier P, Brittain K. An age-old problem: examining the discourses of ageing in HCI and strategies for future research. ACM Trans Comput Hum Interact 2015 Mar 04;22(1):1-27. [doi: 10.1145/2696867]
34. Heponiemi T, Virtanen L, Kaihlanen AM, Kainiemi E, Koponen P, Koskinen S. Use and changes in the use of the internet for obtaining services among older adults during the COVID-19 pandemic: a longitudinal population-based survey study. New Media Soc 2022 Jun 08;26(6):3146-3167. [doi: 10.1177/14614448221097000]
35. European Commission: Directorate-General for Communications Networks, Content and Technology. 2024 digital decade eHealth indicator study: final report. Publications Office of the European Union. 2024. URL: https://op.europa.eu/en/publication-detail/-/publication/c04f6162-3833-11ef-b441-01aa75ed71a1/language-en [accessed 2024-08-13]
36. Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders SM. Methods for human-centered eHealth development: narrative scoping review. J Med Internet Res 2022 Jan 27;24(1):e31858 [FREE Full text] [doi: 10.2196/31858] [Medline: 35084359]
37. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
38. Kujala S, Ammenwerth E, Kolanen H, Ervast M. Applying and extending the FITT framework to identify the challenges and opportunities of successful eHealth services for patient self-management: qualitative interview study. J Med Internet Res 2020 Aug 12;22(8):e17696 [FREE Full text] [doi: 10.2196/17696] [Medline: 32784175]
39. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
40. Charness N, Boot WR. Aging and information technology use: potential and barriers. Curr Dir Psychol Sci 2009 Oct 01;18(5):253-258. [doi: 10.1111/j.1467-8721.2009.01647.x]
41. Currie M, Philip LJ, Roberts A. Attitudes towards the use and acceptance of eHealth technologies: a case study of older adults living with chronic pain and implications for rural healthcare. BMC Health Serv Res 2015 Apr 16;15:162 [FREE Full text] [doi: 10.1186/s12913-015-0825-0] [Medline: 25888988]
42. Nurkka P, Kujala S, Kemppainen K. Capturing users' perceptions of valuable experience and meaning. J Eng Des 2009 Oct;20(5):449-465. [doi: 10.1080/09544820903158835]

43. Kujala S, Kauppinen M, Rekola S. Bridging the gap between user needs and user requirements. In: Advances in Human-Computer Interaction I. Patras, Greece: Typorama Publications; 2001:45-50.
44. Langefors B. Analysis of user needs. In: Proceedings of the 2nd Conference of the European Cooperation in Informatics. 1978 Presented at: ECI 1978; October 10-12, 1978; Venice, Italy. [doi: 10.1007/3-540-08934-9_71]
45. Peine A, Neven L. The co-constitution of ageing and technology – a model and agenda. Ageing Soc 2020 Jun 03;41(12):2845-2866. [doi: 10.1017/s0144686x20000641]
46. Fronemann N, Peissner M. User experience concept exploration: user needs as a source for innovation. In: Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. 2014 Presented at: NordiCHI '14; October 26-30, 2014; Helsinki, Finland URL: https://dl.acm.org/doi/10.1145/2639189.2641203 [doi: 10.1145/2639189.2641203]
47. Backonja U, Hall AK, Thielke S. Older adults' current and potential uses of information technologies in a changing world. Int J Aging Hum Dev 2014 Dec 01;80(1):41-63 [FREE Full text] [doi: 10.1177/0091415015591109] [Medline: 26215298]
48. Fallman D. Design-oriented human-computer interaction. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2003 Presented at: CHI '03; April 5-10, 2003; Ft. Lauderdale, FL URL: https://doi.org/10.1145/642611.642652 [doi: 10.1145/642611.642652]
49. Sinha G, Shahi R, Shankar M. Human computer interaction. In: Proceedings of the 3rd International Conference on Emerging Trends in Engineering and Technology. 2010 Presented at: ICETET 2010; November 19-21, 2010; Goa, India. [doi: 10.1109/icetet.2010.85]
50. Soraghan CJ, Hermann S, Boyle G. User-centered design and older people: introductory considerations for engaging users. In: Hippe Z, Kulikowski J, Mroczek T, Wtorek J, editors. Human-Computer Systems Interaction: Backgrounds and Applications 3. Cham, Switzerland: Springer; 2014:221-231.
51. Dickinson A, Goodman J, Syme A, Eisma R, Tiwari L, Mival O, et al. Domesticating technology: in-home requirements gathering with frail older people? In: Proceedings 10th International Conference on Human - Computer Interaction HCI. 2003 Presented at: HCI 2003; June 22-27, 2003; Crete, Greece p. 827-831 URL: https://napier-repository.worktribe.com/preview/267243/mival2.pdf
52. Valkonen P, Karisalmi N, Kaipio J, Kujala S. Remote interviews and visual timelines with patients: lessons learned. Stud Health Technol Inform 2021 May 27;281:845-849. [doi: 10.3233/SHTI210298] [Medline: 34042793]
53. Gaver B, Dunne T, Pacenti E. Design: cultural probes. interactions 1999 Jan;6(1):21-29. [doi: 10.1145/291224.291235]
54. Lee A, Sandvei M, Asmussen HC, Skougaard M, Macdonald J, Zavada J, et al. The development of complex digital health solutions: formative evaluation combining different methodologies. JMIR Res Protoc 2018 Jul 16;7(7):e165 [FREE Full text] [doi: 10.2196/resprot.9521] [Medline: 30012548]
55. Van den Bulck SA, Hermens R, Slegers K, Vandenberghe B, Goderis G, Vankrunkelsven P. Designing a patient portal for patient-centered care: cross-sectional survey. J Med Internet Res 2018 Oct 01;20(10):e269 [FREE Full text] [doi: 10.2196/jmir.9497] [Medline: 30287416]
56. Bosch L, Kanis M, Dunn J, Stewart KA, Kröse B. How is the caregiver doing? Capturing caregivers' experiences with a reflective toolkit. JMIR Ment Health 2019 May 28;6(5):e13688 [FREE Full text] [doi: 10.2196/13688] [Medline: 31140435]
57. Danbjørg DB, Villadsen A, Gill E, Rothmann MJ, Clemensen J. Usage of an exercise app in the care for people with osteoarthritis: user-driven exploratory study. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e11 [FREE Full text] [doi: 10.2196/mhealth.7734] [Medline: 29326092]
58. Kujala S, Nurkka P. Sentence completion for evaluating symbolic meaning. Int J Des 2012;6(3) [FREE Full text]
59. Kujala S, Walsh T, Nurkka P, Crisan M. Sentence completion for understanding users and evaluating user experience. Interact Comput 2013 Jul 12;26(3):238-255. [doi: 10.1093/iwc/iwt036]
60. Lallemand C, Mercier E. Optimizing the use of the sentence completion survey technique in user research: a case study on the experience of e-reading. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA. [doi: 10.1145/3491102.3517718]
61. Newell A, Arnott J, Carmichael A, Morgan M. Methodologies for involving older adults in the design process. In: Proceedings of the 4th International Conference on Universal Access in Human-Computer Interaction. 2007 Presented at: UAHCI 2007; July 22-27, 2007; Beijing, China. [doi: /doi.org/10.1007/978-3-540-73279-2_110]
62. Chun YJ, Patterson PE. A usability gap between older adults and younger adults on interface design of an internet-based telemedicine system. Work 2012;41 Suppl 1:349-352. [doi: 10.3233/WOR-2012-0180-349] [Medline: 22316747]
63. Boehner K, Vertesi J, Sengers P, Dourish P. How HCI interprets the probes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2007 Presented at: CHI '07; April 28-May 3, 2007; San Jose, CA p. 1077-1086. [doi: https://doi.org/10.1145/1240624.1240789]
64. Isomursu M, Tähti M, Väinämö S, Kuutti K. Experimental evaluation of five methods for collecting emotions in field settings with mobile applications. Int J Hum Comput Stud 2007 Apr;65(4):404-418. [doi: 10.1016/j.ijhcs.2006.11.007]
65. Ehn M, Eriksson LC, Åkerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018 Feb 01;6(2):e34 [FREE Full text] [doi: 10.2196/mhealth.8345] [Medline: 29391342]

66. Hong MK, Lakshmi U, Do K, Prahalad S, Olson T, Arriaga RI, et al. Using diaries to probe the illness experiences of adolescent patients and parental caregivers. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 229-299. [doi: https://doi.org/10.1145/3313831.3376426]
67. Klausen SH. Understanding older adults' wellbeing from a philosophical perspective. J Happiness Stud 2019 Oct 31;21(7):2629-2648. [doi: 10.1007/s10902-019-00197-5]
68. Munteanu C, Axtell B, Rafih H, Liaqat A, Aly Y. Designing for older adults: overcoming barriers to a supportive, safe, and healthy retirement. In: Agnew J, Mitchell OS, editors. The Disruptive Impact of FinTech on Retirement Systems. Oxford, UK: Oxford University Press; 2019:104-126.
69. Moser A, Korstjens I. Series: practical guidance to qualitative research. Part 3: sampling, data collection and analysis. Eur J Gen Pract 2018 Dec;24(1):9-18 [FREE Full text] [doi: 10.1080/13814788.2017.1375091] [Medline: 29199486]
70. Valkonen P, Kujala S, Hörhammer I, Savolainen K, Helminen RR, Vartia I. Health self-management of older employees: identifying critical peak experiences of a patient portal. FinJeHeW 2023 Apr 20;15(2):114-129. [doi: 10.23996/fjhw.126837]
71. Benson T, Potts HW, Bark P, Bowman C. Development and initial testing of a Health Confidence Score (HCS). BMJ Open Qual 2019;8(2):e000411 [FREE Full text] [doi: 10.1136/bmjoq-2018-000411] [Medline: 31259277]
72. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
73. Eronen J, Paakkari L, Portegijs E, Saajanaho M, Rantanen T. Assessment of health literacy among older Finns. Aging Clin Exp Res 2019 Apr;31(4):549-556 [FREE Full text] [doi: 10.1007/s40520-018-1104-9] [Medline: 30578457]
74. Çerçi S, Cecchinato ME, Vines J. How design researchers interpret probes: understanding the critical intentions of a designerly approach to research. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21; May 8-13, 2021; Yokohama, Japan. [doi: 10.1145/3411764.3445328]
75. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
76. Mayring P. Qualitative content analysis: theoretical foundation, basic procedures and software solution. Social Science Open Access Repository. 2014. URL: https://www.ssoar.info/ssoar/handle/document/39517 [accessed 2024-12-15]
77. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
78. Maher C, Hadfield M, Hutchings M, de Eyto A. Ensuring rigor in qualitative data analysis: a design research approach to coding combining NVivo with traditional material methods. Int J Qual Methods 2018 Jul 10;17(1):1-13. [doi: 10.1177/1609406918786362]
79. Nielsen BK, Nielsen LA, Lasgaard M. Loneliness in emerging adulthood: a qualitative study using cultural probes and in-depth interviews. Curr Res Behav Sci 2024;7:100154. [doi: 10.1016/j.crbeha.2024.100154]
80. Kujala S, Miron-Shatz T. Emotions, experiences and usability in real-life mobile phone use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI '13; April 27-May 2, 2013; Paris, France p. 1061-1070 URL: https://doi.org/10.1145/2470654.2466135 [doi: 10.1145/2470654.2466135]
81. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
82. Jarosz E. What makes life enjoyable at an older age? Experiential wellbeing, daily activities, and satisfaction with life in general. Aging Ment Health 2022 Jun;26(6):1242-1252. [doi: 10.1080/13607863.2021.1916879] [Medline: 33908290]
83. Ryu J, Heo J, Ellis G, Widmer MA. Leisure, eudaimonic behavior, physical activity and well-being of older adults. J Leis Res 2022 Jun 08;53(4):595-614. [doi: 10.1080/00222216.2022.2051115]
84. Zhang J, Umemuro H. Older adults' hedonic and eudaimonic well-being in information and communication technology activities. Gerontechnology 2012;10(4):231-243. [doi: 10.4017/gt.2012.10.4.004.00]
85. Heponiemi T, Kainiemi E, Virtanen L, Saukkonen P, Sainio P, Koponen P, et al. Predicting internet use and digital competence among older adults using performance tests of visual, physical, and cognitive functioning: longitudinal population-based study. J Med Internet Res 2023 May 05;25:e42287 [FREE Full text] [doi: 10.2196/42287] [Medline: 37145836]

Abbreviations

DigiIN: Toward a Socially Inclusive Digital Society: Transforming Service Culture Project
eHEALS: eHealth Literacy Scale
HCS: Health Confidence Score
HLS-EU-Q16: European Health Literacy Survey Questionnaire

Edited by C Jacob; submitted 28.06.23; peer-reviewed by L Heryawan, I Mannheim, S Jalil, L Simonse; comments to author 07.12.23; revised version received 01.02.24; accepted 11.11.24; published 08.01.25.

Please cite as:
Valkonen P, Kujala S, Savolainen K, Helminen RR
Exploring Older Adults' Needs for a Healthy Life and eHealth: Qualitative Interview Study
JMIR Hum Factors 2025;12:e50329
URL: https://humanfactors.jmir.org/2025/1/e50329
doi:10.2196/50329
PMID:

Original Paper

Exploring the Impact of Digital Peer Support Services on Meeting Unmet Needs Within an Employee Assistance Program: Retrospective Cohort Study

Harpreet Nagra[1], PhD; Robert A Mines[2], PhD; Zara Dana[1], PhD

[1]Supportiv, Berkeley, CA, United States

[2]MINES & Associates, Littleton, CO, United States

Corresponding Author:
Harpreet Nagra, PhD
Supportiv
2222 Harold Way
Berkeley, CA, 94704
United States
Phone: 1 800 845 0015
Email: harpreet@supportiv.com

Abstract

Background: The World Health Organization estimates that 1 in 4 people worldwide will experience a mental disorder in their lifetime, highlighting the need for accessible support.

Objective: This study evaluates the integration of digital peer support (DPS) into an employee assistance program (EAP), testing 3 hypotheses: (1) DPS may be associated with changes in EAP counseling utilization within a 5-session model; (2) DPS users experience reduced sadness, loneliness, and stress; and (3) DPS integration generates a positive social return on investment (SROI).

Methods: The study analyzed EAP utilization within a 5-session model using pre-post analysis, sentiment changes during DPS chats via natural language processing models, and SROI outcomes.

Results: Among 587 DPS chats, 432 (73.6%) occurred after business hours, emphasizing the importance of 24/7 availability. A matched cohort analysis (n=72) showed that DPS reduced therapy sessions by 2.07 per participant (P<.001; Cohen d=1.77). Users' messages were evaluated for sentiments of sadness, loneliness, and stress on a 1-10 scale. Significant reductions were observed: loneliness decreased by 55.04% (6.91 to 3.11), sadness by 57.5% (6.84 to 2.91), and stress by 56.57% (6.78 to 2.95). SROI analysis demonstrated value-to-investment ratios of US $1.66 (loneliness), US $2.50 (stress), and US $2.58 (sadness) per dollar invested.

Conclusions: Integrating DPS into EAPs provides significant benefits, including increased access, improved emotional outcomes, and a high SROI, reinforcing its value within emotional health support ecosystems.

(JMIR Hum Factors 2025;12:e68221) doi:10.2196/68221

KEYWORDS

digital peer support; peer support; EAPs; cost-effectiveness; SROI

Introduction

Background

The World Health Organization estimates that 1 in 4 people globally will experience mental disorders in their lifetime [1], stressing the need for a readily available treatment ecosystem. Relatedly, the increasing popularity of remote work post–COVID-19, along with work- and workplace-related stress—stemming from long hours, high workloads, an inability to disconnect, and, in particular, isolation—has contributed to a rise in mental health problems such as anxiety and depression, musculoskeletal pain, sleep deprivation, and stress [2]. These stressors impact not only employees' well-being but also presenteeism, absenteeism, and work-life balance [2-4]. Burnout, in particular, arises from chronic stress characterized by emotional, physical, and mental exhaustion, leading to reduced productivity, increased absenteeism, and higher turnover rates [3].

While the Centers for Disease Control and Prevention's National Center for Health Statistics reports a notable increase in treatment-seeking behaviors between 2019 (19.2%) and 2021 (21.6%), a significant majority of working adults may still suffer alone [5]. The proportion of working adults aged 18-44 years receiving mental health treatment grew significantly from 18.5% in 2019 to 23.2% in 2021 [5], with women (23.8% to 28.6%, respectively) continuing to outpace men (13.1% to 17.8%, respectively) [5]. By 2021, adults aged 45-64 years (21.2%) and those aged 65 years and over (18.9%) ranked second and third, respectively, in seeking treatment [5].

Employee Assistance Programs

Mental and emotional health–focused employee assistance programs (EAPs), which are confidential, voluntary, organization-sponsored, and low-cost services for employees, offer personal or work-related counseling and support. The purpose of EAPs is to help employees manage stress, enhance problem-solving abilities, and, in turn, improve work productivity, reduce absenteeism, and promote overall well-being [6]. Counseling-based EAPs provide psychoeducation, teach coping strategies, and assist users in setting achievable personal goals that build confidence and self-efficacy [7-12]. Additionally, EAPs offer training in communication skills, such as assertiveness, to improve relationships and help users navigate challenges effectively. Patient activation, which refers to an individual's knowledge, skills, confidence, and willingness to manage their health and well-being [7,8], is further supported by connecting users to resources and support networks, including health care providers and peer support groups. Finally, EAPs empower users with self-management tools, such as digital apps and resource libraries, that allow them to track their progress and independently manage their well-being [12]. Together, these strategies foster self-reliance, enhancing users' ability to manage their mental health and overall well-being [7]. Depending on the EAP, a significant return on investment (ROI) may be observed, with some organizations seeing US $3-US $10 for every dollar spent [13], largely due to positive impacts on absenteeism, productivity, and employee turnover and retention [6,13]. EAPs are well-established assets to organizations, delivering emotional benefits to employees and productivity and financial benefits to employers [6,13].

Digital Peer Support

While counseling-based EAPs are known for their advantageous returns, not every emotional concern requires counseling support. A tiered support system for individuals who are mentally well but experiencing lower-acuity emotional distress may be necessary. Simultaneously, the US health care system currently faces a shortage of mental health professionals, limiting access to adequate support [14,15]. In 2021, 5930 areas were designated as health professional shortage areas, leaving 129.6 million people without access to affordable and accessible care [15]. President Biden's 2022 White House brief highlighted this issue and called for efforts to "strengthen system capacity," including the increased use of peer support specialists and paraprofessionals [16].

Peer support interventions involve individuals with similar lived experiences offering mutual support [17]. While expert-based professional support may evoke feelings of shame, failure, and mental health stigma, peer support can normalize emotional struggles and foster connection, validation, and empathy through shared experiences [18]. This "mutual empowerment" from peer-driven interventions grants individuals greater autonomy and competency [18]. However, some Americans may not have access to digital peer services due to a lack of universal internet access and low health literacy. Approximately 24 million people in the United States lack reliable broadband service or high-speed internet [19]. Historically, lower rates of high-speed internet use have been observed in households where the owner is 65 years or older, Hispanic, African American, American Indian, or Alaska Native [19]. Furthermore, in 2021, only about 5 in 10 households with incomes below US $25K used high-speed internet [19]. Additionally, approximately 90 million American adults struggle to understand and navigate complex health- and text-based tasks accurately and consistently, exhibiting low health literacy [20]. By addressing these disparities, digital peer support (DPS) services could become even more inclusive by expanding support to those currently excluded from traditional services.

Self-determination theory, as articulated by Ryan and Deci [21], offers insight into key features that make peer support impactful. The theory states that self-determination and emotional growth are driven by 3 core elements—autonomy, competence, and relatedness [21,22]. Autonomy refers to the ability to self-direct and control one's actions, competence involves mastering new skills, and relatedness is about forming meaningful connections with others [21,22]. Peer support provides a nonjudgmental, nonstigmatizing space that fosters autonomy, as there is no professional expert directing treatment. Competence is strengthened through the normalization of mistakes shared via lived experiences, as well as the application and practice of coping strategies. Relatedness is reinforced through the verbal and emotional connections formed with similar peers [18]. Research has consistently shown that relatedness in supportive relationships contributes significantly to emotional regulation, motivation, and the overall success of interventions. By reinforcing relatedness through peer-to-peer interactions, peer support programs address a fundamental psychological need that complements autonomy and competence, further validating their essential role in the model [17,18]. The interrelatedness of these dynamic factors is believed to enhance mental health and well-being [17-22]. The introduction of DPS platforms extends these benefits by offering accessible, mutually supportive communities to help alleviate feelings of loneliness and isolation [23]. Numerous studies have demonstrated the effectiveness of these digital interventions in addressing mental health needs [23-30].

Confirmed benefits of peer support interventions are evident through self-determination theory and digital platforms [23-30]; however, a significant gap remains in understanding the ROI, particularly when integrated into counseling-based EAPs. Studies suggest that peer support may lead to direct health care savings by equipping individuals with coping mechanisms and providing emotional support, which, in turn, reduces the risk

of crises and costly subsequent interventions (eg, inpatient hospitalization, emergency room use) [31-33]. Specifically, current research does not adequately address how incorporating 24/7 DPS into EAPs translates into measurable cost-effectiveness for organizations. Traditionally, cost-effectiveness and cost-benefit analyses are used to assess the value for money of health and social interventions, but the value derived from participating in 24/7 DPS can be subtle and difficult to measure [34]. As a result, there is a paucity of research on the broader social, economic, and environmental value. This gap underscores the need for a focused study to evaluate the social return on investment (SROI) of integrating 24/7 DPS into a counseling-based EAP, assessing both the economic impact and the efficiency of these interventions.

Social Return On Investment

SROI has become a recognized method for measuring the impact, outcomes, and value created by social-emotional–focused organizations [35,36]. Broadly, SROI uses a mixed-method design to assess the value of an intervention relative to the cost of enabling it. Beyond financial metrics, which may assign limited value to a program, SROI captures social, environmental, and economic elements to generate a comprehensive "social value."

Study Aim and Hypotheses

For this study, a 24/7 US-based DPS service was integrated with a US-based counseling-based EAP firm to explore potential changes in utilization, sentiment, and the SROI of the 24/7 DPS for the EAP's clientele. The following study hypotheses were examined: (1) The introduction of the 24/7 DPS service may lead to changes in the utilization of EAP counseling services within a 5-session model; (2) participants who utilize the DPS service will experience significant changes in sentiment (eg, reduced sadness, loneliness, and stress) over the course of their engagement with the service; and (3) the integration of DPS into the EAP will produce a positive SROI, reflecting the added value of peer support services for EAP clientele. Through these key objectives, the study aims to assess the effects of DPS on EAP utilization, emotional well-being, and the social and economic value added to the EAP.

Methods

Pre- and Post-DPS Integration: EAP Services and Implementation

Using aggregated participant data, pre- and postanalyses of adding DPS to EAP services were conducted. Pre-DPS was defined as the routine care provided within the EAP, including an initial intake, screening for clinical concerns, and offering service options such as counseling, legal resources, psychoeducation, and coaching. Before the introduction of DPS, participants either accessed the EAP provider's website using a company code to explore available services or completed an initial online or phone intake. They were screened for clinical concerns, including substance use and suicidal or homicidal ideation, by licensed clinicians or clinically supervised paraprofessionals. The EAP offers a comprehensive suite of behavioral health and wellness services designed to support mental health, work-life balance, and overall well-being. These offerings include free and confidential counseling, legal and financial benefits (including Medicare and Social Security consultation), wellness coaching, work-life referral services, and an online resource library. Additional services include virtual reality programs, as well as parenting, life, career, and work performance coaching. The EAP provides managed care 24/7, with crisis support delivered by licensed mental health professionals.

An anonymous US-based DPS service that provided 24/7/365 moderated, synchronous peer group chats was integrated into the US-based EAP's client support ecosystem. The DPS service utilized artificial intelligence (AI)–driven natural language processing to match users with peers facing similar issues in small groups, each facilitated by a trained human moderator who ensured a safe environment and directed users to professional services in crises. The DPS model allows for unlimited sessions, in both duration and frequency, while the EAP's 5-session model was selected for further analysis due to the prevalence of 5-session models over other session formats.

Participants

DPS: Peer Support Users

Nonidentifying participant data from the DPS provider, collected between June 2023 and May 2024, were used to evaluate the research objectives. DPS service participants had access to unlimited peer support chats and were included in the data analysis if they utilized the chat service at least once.

EAP: Counseling Service Participants

Inclusion Criteria

Pre-DPS (June 2022 to May 2023) and post-DPS (June 2023 to May 2024) participants were selected based on having active access to the 5-session EAP model. For the analysis of total utilization, no demographic factors were matched between the pre- and post-DPS groups. However, for the utilization analysis of the 5-session EAP model and to ensure comparable groups, participants in both the pre-DPS and post-DPS cohorts were matched based on demographic and emotional concern variables. Matching was conducted using propensity score matching, adjusting for potential confounders such as age, gender, and emotional concerns (eg, anxiety or depression). Participants who had utilized the full 5-session model in both groups were selected, with efforts made to capture diverse client trajectories by including multiple participants matched on these same variables. EAP participants were employed across a wide range of sectors, including health care, education, professional services, manufacturing, construction, building trades, nonprofits, technology, service industries (eg, restaurants), unions, mining, government, and others.

Cohort Exclusion Criteria

Participants were excluded from the analysis if they had missing demographic data (age or gender) or if they exceeded the authorized number of counseling sessions (eg, >5 sessions in the pre-DPS group). These exclusions were necessary to control for confounding variables, as participants requiring more intensive mental health support (beyond 5 sessions) may have

had significantly different needs compared with those who adhered to the 5-session model. This approach ensured that the study focused on participants whose engagement with EAP services reflected typical utilization patterns.

Peer Support Moderators

Peer support moderators (PSMs) are human moderators trained in digital safety monitoring who provide real-time, text-based support. PSMs receive up to 164 hours of training within the first 90 days of hire, focusing on engaging in digital conversations, facilitating synchronous group chats, and managing the psychological safety of chat users by monitoring for and removing trolls, as well as enacting safety protocols, such as referring users who may be in active crisis to appropriate services. Additionally, PSMs receive 36 hours of ongoing asynchronous and synchronous clinical supervision and consultation per quarter.

Procedures

Overview

This retrospective study evaluated the differences in the EAP's services before and after DPS to assess the programmatic impact and added social value.

Data Variables

The primary data sources for this study were aggregated engagement variables, such as chat time, number of chats, age, gender, user struggles or presenting concerns, and national-level cost-of-service value data.

Data Extraction and Management

The data extraction process involved retrieving DPS chat data, which included digital chat records, user interactions with the platform, and usage patterns. All collected data were deidentified and anonymized to ensure participant privacy and confidentiality. Data management procedures adhered to institutional and regulatory guidelines, ensuring data security and integrity, with access restricted to authorized personnel. Measures were in place to maintain data security, integrity, and confidentiality throughout the analysis process.

Statistical Analysis

Analysis and Group Comparability

Descriptive statistics were used to summarize participant characteristics and their engagement with DPS and EAP services. To assess differences in service utilization and sentiment scores between pre-DPS and post-DPS groups, paired *t* tests (2-tailed) were conducted for continuous data (eg, session frequency), and chi-square tests were used for categorical data (eg, gender, emotional concern). As DPS is anonymous, it was not possible to directly track which EAP participants used the service. Using volunteered age and gender data from DPS participants as a starting point, propensity score matching was applied to control for potential confounders such as age, gender, and emotional concern, ensuring group comparability before DPS integration.

Mood Categorization

Clinical counseling experts reviewed the emotional concerns reported by DPS participants, categorizing them into mood categories such as depression, anxiety, adjustment, and relational concerns to align DPS participants with the presenting emotional concerns of EAP participants. The counseling experts were 2 licensed psychologists with doctoral degrees in counseling psychology and a combined 50 years of experience in the field. The clinical interpretation process, guided by the Biopsychosocial Model [37], involved categorizing participants' emotional struggles into mood categories. This model integrates biological, psychological, and social dimensions to provide a comprehensive understanding of mental health. For example, participants' reports of feeling "trapped" or "isolated" were examined through a psychological lens to identify depressive thought patterns, while work-related stress (eg, "I hate my boss") was contextualized within the social environment to classify relational concerns [38]. By incorporating these factors, the model enabled clinical experts to holistically interpret participants' struggles, ensuring that the categorizations captured the complexity of their emotional experiences. The categorization process involved initially reviewing the raw data from participants' reported "struggles" and presenting concerns in therapy individually, followed by cross-validation between the 2 experts to ensure consistency. This collaborative step helped reconcile differences in categorization and address overlapping cases. It is important to note that neither the text-based struggle data from DPS nor the presenting concerns from EAP represented formal diagnoses; rather, they reflected participants' personal interpretations of their symptoms. Keywords such as *helpless* or *sadness* were flagged for depression-related presenting concerns, while terms related to *fear, excessive worry,* or *panic* were matched with anxiety-focused concerns. In cases where symptoms overlapped (eg, depression and anxiety), the experts aligned their categorizations based on the primary presenting concern reflected in the participant's statement. For example, if a participant's statement indicated feelings of hopelessness and despair alongside anxious thoughts, the struggle was categorized as depression. Alternatively, if worry or nervousness predominated, it was categorized as anxiety. In scenarios where a specific relationship was mentioned alongside anxious or depressive feelings, the struggle was categorized primarily as a relational concern. Symptom overlap was resolved by categorizing the statement according to the dominant theme present in the participant's experience. Table 1 provides examples of these interpretations.

Table 1. Clinical interpretation of comparable emotional concerns shared on digital peer support service with employee assistance program service.

Participant's struggle	Presenting concern[a]
"I am feeling pretty suicidal and trapped. I am new to my work from home job and I need help, resources, a plan something or this will only get worse. I am trapped with no transportation, no friends or any meaningful relationships in life. Just a work schedule and I'm isolated."	Depression
"I hate my job and my boss"	Relational concern
"I've been feeling a lot of anxiety around work and my personal life"	Anxiety/stress
"Alcohol addiction"	Substance use

[a]These are not diagnoses, they are layperson understandings.

Sentiment Analysis

Peer-to-peer conversations were analyzed to produce a quantitative measure of emotional change. Each participant's chat session, referred to as a "user-chat-session" or simply "user-chat," was treated as an independent unit of analysis, regardless of whether the same participant took part in multiple sessions. As the DPS service is anonymous, each user and their experience in a given chat session were considered separately, without combining data from different sessions for the same user. This approach ensures that each user's experience is analyzed in isolation, providing a more accurate, user-centric view of emotional changes.

Each user's message was analyzed in the context of preceding messages from other users to capture their emotional journey throughout the chat session. Context was defined as the concatenation of all preceding messages from other users in the same chat, starting from the latest message of the user of interest. A user's initial struggle and first message had no prior context and were evaluated with an empty context. For example, in the following conversation:

userA: message1

userB: message2

moderator: message3

userA: message4

userB: message5

This example assumes that message1 and message2 are the first messages of user A and user B, respectively, in the chat session. User A's emotional journey was quantified by evaluating message1 with an empty context and message4 in the context of message2 + message3. Similarly, for user B, message2 was evaluated without context, and message5 was evaluated in the context of message3 + message4.

Each user's message, along with its corresponding context, was evaluated for sentiments of sadness, loneliness, and stress using a scale of 1-10, where 1 indicates low intensity, 5 indicates moderate intensity, and 10 indicates high intensity of the sentiment. The analysis focused on chat sessions where participants exhibited at least moderate sentiment intensity (score≥5) from the beginning.

A third-party, public natural language processing model from OpenAI, called GPT-4, was used to assign sentiment scores to each user message and its preceding context [39]. A few-shot learning approach was applied, where the model was provided with 10 examples representing different sentiment intensities to calibrate its interpretations [40]. GPT-4 demonstrates an advanced capacity to interpret emotion, surpassing benchmarks from the general population and offering highly congruent performance in emotion interpretation [41]. See Table 2 for examples provided to GPT-4 to predict loneliness levels.

Sentiment analysis evaluated each user's chat messages in temporal order. The quantified emotions were then interpolated and extrapolated to generate a collective sentiment trend across peer-to-peer chats of varying lengths. A one-sided Mann-Whitney *U* test was used to compare emotion scores at the start and end of the conversations.

Table 2. Examples provided to GPT-4 to interpret loneliness intensities in a few-shot manner.

Intensities	Loneliness
1=low	"Take care everyone. Thanks for the great talk"
5=moderate	"I went on Monday and today. After today's class is when I felt worse as I chatted with a lady in class who told me that she got really depressed when she retired."
10=high	"I just always feel totally alone. Like no one understands what is going on in my head"

SROI Methodology

Historically, the SROI methodology has relied on qualitative methods to identify outcome themes most relevant to participants, followed by quantitative approaches to determine monetary proxies for the identified outcomes and related values [41].

With the advent of natural language processing models, significantly larger volumes of text-based data can now be analyzed, allowing qualitative data to be graphically represented. This study utilized text-based data from peer group participants—including open-ended questions, affirmations, reflective questions, and summaries—to examine changes in sentiments such as sadness, loneliness, and stress within

peer-support chats [42]. The SROI methodology was applied to assess the social value created by integrating DPS into the EAP. Qualitative data from peer-to-peer conversations, including open-ended questions, affirmations, and summaries, were analyzed to understand sentiment changes and emotional shifts within the peer group chats. Quantitative methods were used to calculate a monetary proxy for the emotional outcomes. The Sopact SROI calculator, available on the Sopact website [41], was utilized to determine the social value generated by the DPS service, incorporating factors such as enhanced emotional well-being and cost savings from reduced therapy usage. Given the novelty of DPS and its integration into the EAP, a small sample size was anticipated. This limitation was addressed through rigorous matching techniques and statistical analyses to ensure the findings remained valid and robust despite the smaller sample size. Moreover, it is acknowledged that some participants may have preferred individual counseling services over DPS, potentially introducing selection bias among those who opted for the DPS model.

Ethical Considerations

Informed consent was obtained from all participants before their involvement in the study. The consent process included clear explanations of the research's purpose, data collection methods, and the measures implemented to ensure participant privacy and confidentiality. All data collected were deidentified to safeguard participants' anonymity. The study complied with all institutional and regulatory guidelines related to data protection and participant privacy. To ensure participant safety within the DPS, real-time moderation was provided by trained human moderators specializing in digital safety monitoring. Furthermore, the Pearl Institutional Review Board reviewed and approved the study (approval ID 2024-0442).

Results

Total DPS Utilization

- Study hypothesis 1: The introduction of the 24/7 DPS service may be associated with utilization changes of EAP counseling services within a 5-session model, specifically among participants who engage with DPS.

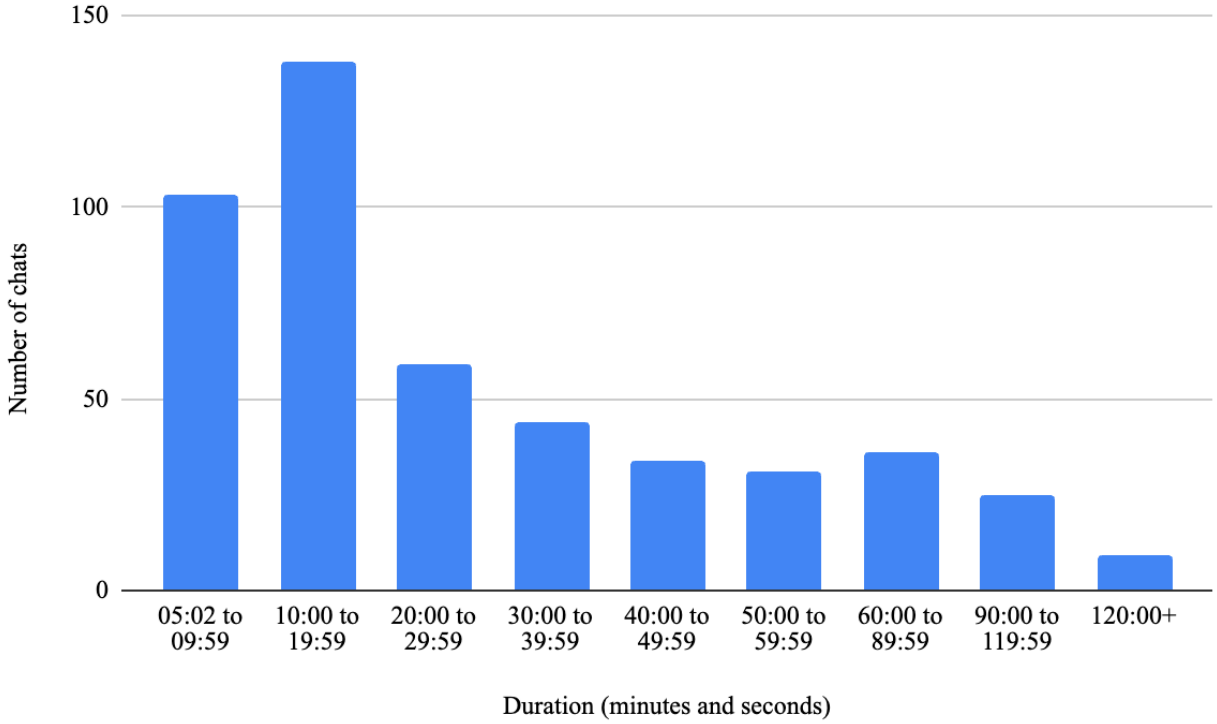
Out of 587 total chats, 432 took place between 5 PM and 8 AM, hours when most therapy offices are typically closed. Chats lasting less than 5 minutes were excluded from the analysis, as the initial 5 minutes are generally used for introductions and setting expectations [30]. Consequently, 107 chats were omitted. A summary of the data is provided in Table 3.

When categorized by duration, the data revealed that 103 chats lasted between 0 and 9 minutes 59 seconds, 138 chats between 10 and 19 minutes and 59 seconds, and 59 chats between 20 and 29 minutes and 59 seconds. Additionally, 44 chats lasted between 30 and 39 minutes and 59 seconds, 34 chats between 40 and 49 minutes and 59 seconds, and 31 chats between 50 and 59 minutes and 59 seconds. While longer chats were less common, they occurred consistently: 36 chats lasted between 60 and 89 minutes and 59 seconds, 25 chats between 90 and 119 minutes and 59 seconds, and 9 chats exceeded 120 minutes. Refer to Figure 1 for further details.

Table 3. Peer support program utilization data.

Metric	Value
Total chats, n	480
Total messages, n	7219
Total duration, hours:minutes:seconds	254:17:57
Average duration, hours:minutes:seconds	0:31:51
Shortest duration, hours:minutes:seconds	0:05:02
Longest duration, hours:minutes:seconds	3:19:11

Figure 1. Chat duration distribution.



Pre-Post Analysis

Matched Cohort Analysis

To evaluate the utilization impact of DPS on the EAP's 5-session model, semimatched cohorts were created for the periods before (June 2022 to May 2023) and after DPS implementation (June 2023 to May 2024). Because of the anonymous nature of the DPS service, only participants who voluntarily provided their age and gender were included in the matching process for analysis in the EAP's pre- and post-DPS cohorts. The matching variables comprised age, gender, and presenting emotional concerns.

Sample

Data on presenting emotional concerns were readily available based on participants' responses to the question, "What's your struggle?"—the sole query asked before matching them with the DPS service. A total of 45 unique users provided both their age and gender information voluntarily. These users ranged in age from 18 to 51 years, with a mean age of 34.97 (SD 9.71) years. Gender distribution was as follows: 18 (40%) males, 26 (58%) females, and 1 (2%) nonbinary individual.

The availability of participant age, gender, and emotional concern data from the DPS service enabled the creation of a matched pre-DPS EAP cohort. The frequency of emotional concerns in the peer support program sample is presented in Table 4.

Table 4. Emotional concern frequency in the digital peer support sample (n=45).

Emotional concern type	Value, *n* (%)
Adjustment	1 (2)
Depression	18 (40)
Anxiety	8 (18)
Posttraumatic stress disorder	2 (4)
Relational concern	14 (31)
Substance use	2 (4)

Statistical Analysis

Using age and gender data from the DPS service participant pool, a total sample of 72 EAP participants—36 pre-DPS and 36 post-DPS—was analyzed for differences. The ages of EAP participants ranged from 18 to 51 years, with a mean age of 33.86 (SD 8.18) years. Females comprised the majority of participants (n=68, 94%), while males accounted for a smaller proportion (n=4, 6%) in both the pre- and post-DPS cohorts.

To assess the impact of the DPS addition, a paired samples *t* test (2-tailed) was performed, comparing the number of therapy

sessions utilized by participants pre- and post-DPS within the matched cohort. The analysis revealed a mean difference of –2.07 (SD 1.77) sessions (t_{71}=9.92; P<.001), indicating a significant reduction in the number of therapy sessions used after the intervention. These findings suggest that the intervention may have been associated with a decrease in therapy session utilization among the sampled participants.

Effect Size

In addition to the *t* test, Cohen *d* was calculated to determine the effect size of the intervention. Based on Cohen guidelines, effect sizes are classified as small (0.2), medium (0.5), and large (0.8). The effect size in this study was 1.77, which is considered large. This suggests that the DPS service had a substantial impact on reducing the number of therapy sessions utilized by participants.

The CI for the mean difference in therapy session utilization was calculated to be between –1.47 and –0.87 at the 95% CI level. This range further supports the reliability of the observed reduction in session utilization, providing evidence that the DPS service may have effectively decreased the number of therapy sessions used by participants in the small sample group.

- Study hypothesis 2: Participants who utilize the DPS service will experience significant changes in sentiment (eg, reduced sadness, loneliness, and stress) throughout their engagement with the service.

Sentiment Change

From the DPS service's sampled cohort (n=45), sentiment changes for sadness, loneliness, and stress were investigated for each user chat. The results are presented in Figures 2-4, showing the changes in sentiment for these 3 emotions. All 3 emotions demonstrated statistically significant reductions (sadness, loneliness, and stress: P<.001) in sentiment intensity during the chat sessions. The sentiment intensity levels are categorized as follows: 1-4 as "low," 4-7 as "moderate," and 7-10 as "high."

Figure 2. Significant (P<.001) changes in sadness sentiment during chats.

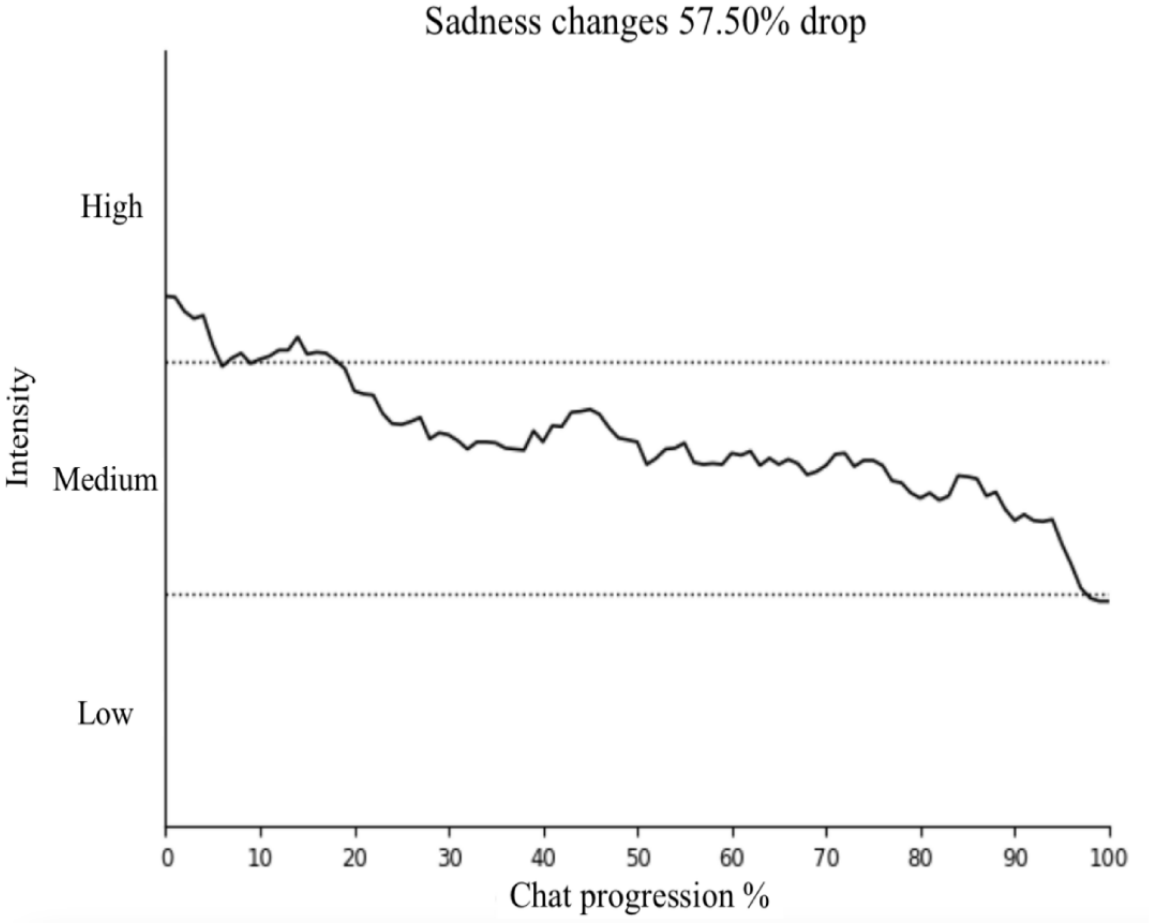


Figure 3. Significant (*P*<.001) changes in loneliness sentiment during chats.

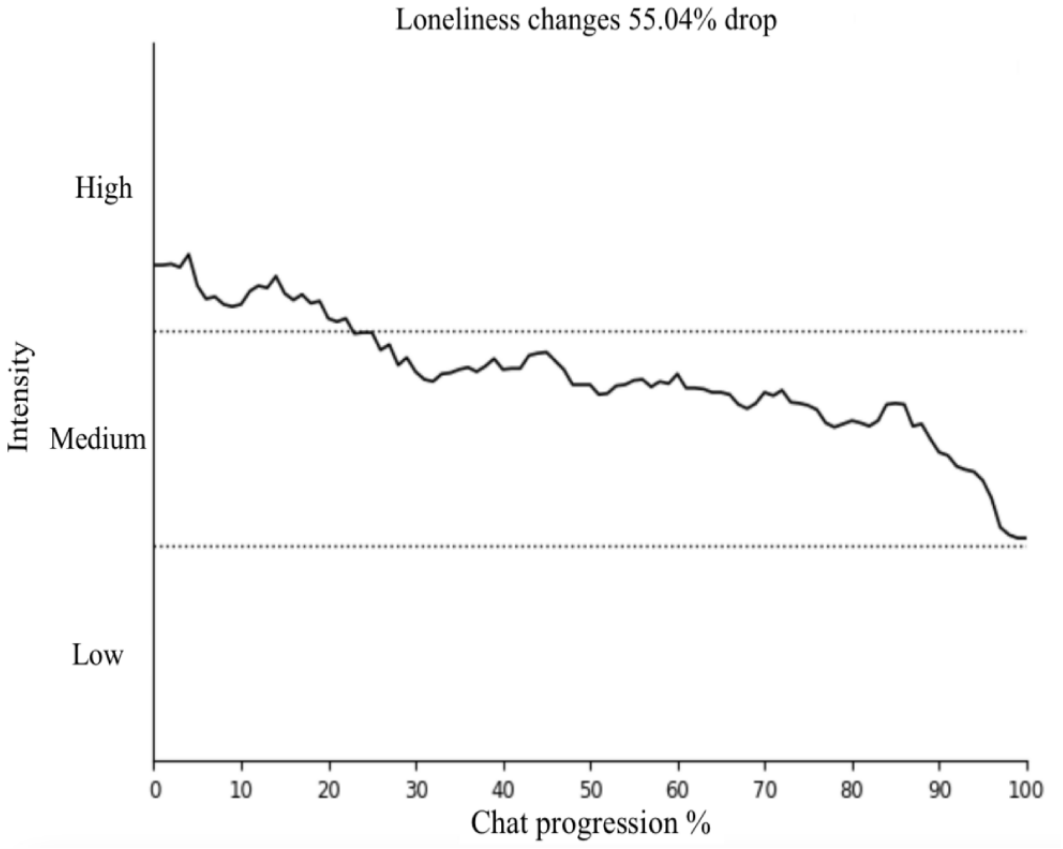
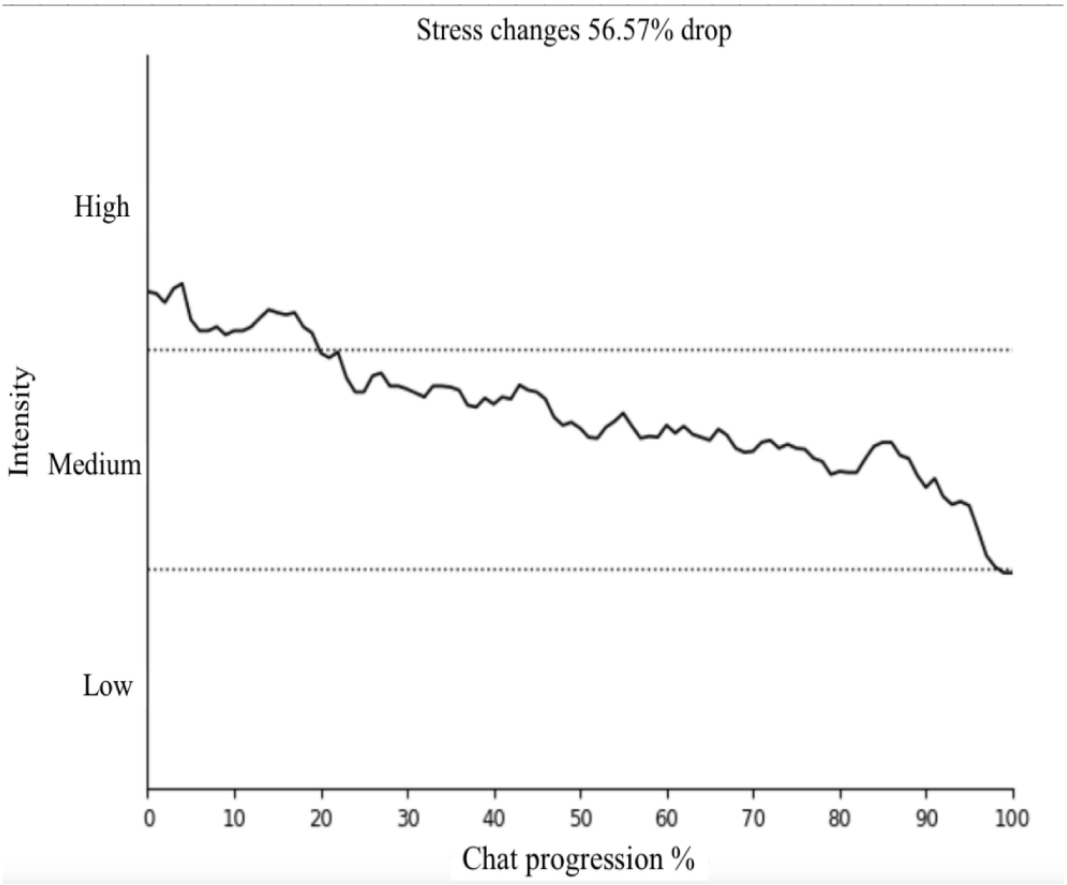


Figure 4. Significant (*P*<.001) changes in stress sentiment during chats.



Social Return on Investment Analysis

SROI assigns financial proxies to quantify the value of each identified sentiment change, enabling the attribution of a market price when no direct market value exists [41]. The proxies were carefully selected based on their relevance and supported by cited research studies to ensure the credibility of the assigned market prices. Table 5 presents the financial proxies used for each sentiment.

- Study hypothesis 3: The integration of DPS into the EAP will generate a positive SROI, reflecting the added value of peer support services for EAP clientele.

Table 5. Financial proxies for each sentiment.

Factors in which digital peer support users experience a reduction	Financial proxy	Cost
Sadness	Average unit cost of treating someone with depression (National Institutes of Health)	• PPPY[a]: US $10,074 (SD $25,694) • PPPM[b,c]: US $839.50
Loneliness and isolation	Average unit cost of treating someone with objective isolation and loneliness (National Institutes of Health)	• PPPY: US $1643 in Medicare spending (objective isolation) • PPPM: US $136.92
Stress (eg, generalized anxiety)	Average unit cost of treating someone with stress	• PPPY: US $6475 • PPPM: US $539.58

[a]PPPY: per person per year.

[b]PPPM: per person per month.

[c]PPPM calculations were made by dividing the PPPY value by 12 months.

SROI Analysis

The primary objective of this SROI analysis is to calculate and present a ratio that compares the cost of investment (US $) in an intervention with the total social, environmental, and economic value (US $) it generates. To prevent overestimation, the SROI methodology recommends certain considerations (see Table 6) [41].

The SROI analysis demonstrated that the DPS program generates positive social value, complementing overall treatment. The SROI values ranged from US $1.66 (loneliness) to US $2.50 (stress) to US $2.58 (sadness) for every dollar invested. Detailed calculations for the impact ratio are available in Multimedia Appendix 1.

Table 6. Factors considered for calculating the social return on investment impact ratio.

Factor	Description
Deadweight	A measure to describe the amount of an outcome that would have happened anyway, even if the DPS[a] service had not been offered. For example, the participants experiencing reductions in momentary loneliness may have a deadweight value of 10% from simply joining the DPS chat, but without further intervention from the peer support moderator or other peers.
Displacement	An assessment of what activities or services are displaced by the presence of the digital peer group. There was no evidence of displacement as, before the digital support option was added, there were no after-hours DPS chat services for employee assistance program members to access.
Attribution	A measure to consider how much of an identified theme is a result of the group studied or is influenced by external factors, for example, if participants also attend another service, such as counseling, which reduces their sense of chronic sadness, a conservative estimate of 20% of the identified momentary sentiment change may be attributed to the group studied.
Dropoff	The percentage of the outcome that decreases after the intervention is complete. For example, a conservative percentage of 10% re-emergence of stress may be expected after the user ends their DPS chat due to a return to their stressful environment.

[a]DPS: digital peer support.

Discussion

Principal Findings

The study examined the impact of integrating a DPS service within an EAP on utilization rates, sentiment change, and therapy session use over 1 year. Additionally, it assessed the social, environmental, and economic impact of the DPS service by estimating an SROI value. Results indicated that DPS service utilization primarily occurred after hours. In the matched sample, participants used fewer therapy sessions post-DPS integration. The SROI ranged from US $1.66 to US $2.50 to US $2.58 for every dollar invested for loneliness, stress, and sadness sentiments, respectively. These findings suggest that combining DPS with traditional EAP services may provide noncrisis

emotional support after hours, potentially reducing EAP therapy session use for lower-acuity emotional concerns. This, in turn, could allow EAPs to allocate funds toward additional ancillary support services and enable employers to benefit from a positive SROI.

The study's findings—that participants experience significant improvements in sentiment, such as reduced sadness, loneliness, and stress—align with existing literature on patient activation [7,8]. Research indicates that when individuals are activated—meaning they gain confidence and skills in managing their health—they tend to experience improvements in emotional well-being [8]. By providing peer support that fosters autonomy and self-efficacy, participants can develop coping strategies that reduce distress and promote mental health [43]. This aligns with findings that patient activation leads to better emotional outcomes through the development of personal coping mechanisms and self-management [8]. The positive SROI from integrating DPS into EAP services is also consistent with the literature on patient activation. Activated patients, who feel more in control of their health, contribute to more efficient resource utilization and improved outcomes [44]. The integration of peer support services such as DPS has been shown to enhance engagement with formal mental health services, creating value not only for individuals but also for the broader service system [44]. This supports the idea that empowering individuals through peer support can yield both personal and systemic benefits, ultimately leading to a positive SROI.

The study may also contribute to the quality of life literature, which is commonly used in cost-effectiveness studies. Health-related quality of life (HRQoL) refers to a person's physical, emotional, and social well-being as influenced by a health intervention [45]. While this study did not include standardized HRQoL assessments (eg, 36-item Short Form [SF-36] or EQ-5D) due to real-world data collection constraints and the anonymous nature of the DPS service, sentiment changes and therapy utilization could serve as HRQoL proxies [46]. Significant reductions in sadness, loneliness, and stress during DPS chats suggest improvements in emotional well-being, a core component of HRQoL. The reduction in momentary negative sentiments may indicate that participants used DPS as a coping mechanism for emotional stability. Additionally, the suggested decrease in therapy sessions implies that participants may have achieved better emotional regulation or relief through DPS, ultimately impacting their overall quality of life. Relatedly, the high utilization of DPS after hours addresses a gap in traditional mental health services and may contribute to participants' perceived control over their emotional health. Such improvements in HRQoL proxies could translate into better job performance [47] and employee satisfaction [48], while also serving as a valuable complement to therapy. This approach allows mental health providers to focus on higher-acuity cases while DPS services address lower-acuity concerns—key objectives for most EAP organizations [48]. Finally, the calculated SROI highlights tangible benefits for emotional health (eg, sadness, loneliness, stress), further demonstrating the intervention's effectiveness in enhancing HRQoL's socioemotional components.

The study's findings further align with existing literature in several key areas. The use of AI for matching participants with peers who have shared lived experiences, as well as for sentiment analysis, supports growing evidence that AI can enhance the personalization and efficiency of life struggle and nonmental health services [49-52] This study offers several novel contributions to the literature on mental health interventions and EAPs. It is among the first to demonstrate that integrating a DPS service with EAPs may enhance after-hours support, reduce the utilization of EAP therapy sessions—allowing EAPs to allocate resources more effectively for higher-acuity emotional concerns—and provide supplementary SROI. DPS services are unique within the EAP environment and may be particularly valuable for EAP participants who are not yet ready to confront the stigma or commitment often associated with therapy [53] but still benefit from peer support. By leveraging AI-driven natural language processing to match users based on their struggles and conducting sentiment analysis on chat narratives, this study highlights an innovative use of technology to enhance support for stress and life challenges. Conducted in a naturalistic setting with a diverse sample, the study's design strengthens ecological validity, demonstrating the practical feasibility and effectiveness of DPS in real-world organizational environments.

Study Limitations

This study faced several potential threats to both internal and external validity and implemented strategies to mitigate these issues. One threat to internal validity was the influence of external events (history) during the study period (June 2023 to May 2024) [54,55]. To address this, a comparison group from the previous year (June 2022 to May 2023) was used, and quarterly data monitoring helped identify and adjust for significant external changes [55]. Matched samples based on additional demographics were not possible due to the anonymity of DPS users. Future research using alternative study designs with nonanonymized data may be necessary.

Maturation effects, referring to the natural changes in participants' mental health over time, were addressed using a cohort design with propensity score matching. This approach ensured that participants in both the intervention and control groups were similar in age, gender, and presenting concerns [56]. By balancing the groups, it minimized the impact of maturation on the results [56]. Additionally, selection bias was mitigated through propensity score matching, which created comparable groups by matching participants in the pre-DPS and post-DPS periods based on key demographics and presenting concerns [57].

Experimental mortality, or dropout, was another concern. To address this, the study included multiple participants matched on the same variables, ensuring that the analysis could proceed even if some participants dropped out [57]. A diverse sample across age and gender was included to reduce the impact of selection biases on generalizability. Additionally, conducting the study in a naturalistic setting, where participants used the DPS and EAP services as they would in real life, helped minimize the reactive effects of experimental arrangements, enhancing ecological validity [55]. Furthermore, the availability

and normalization of the EAP program may have influenced participant engagement. EAP programs are voluntary for companies in the United States, Australia, Canada, and England, whereas Nordic countries have legal mandates requiring mandatory EAPs [58]. This distinction is relevant from the perspective of patient activation. Voluntary EAPs may require high awareness and active promotion to engage participants and may primarily attract individuals with higher activation or those who are more self-directed. As a result, individuals with lower activation or those in the early stages of their mental health concerns may remain underserved [58]. Additionally, factors such as program promotion and mental health stigma could contribute to uneven access across the workforce. By contrast, mandatory EAPs may help normalize access to mental health support, encourage early and preventive engagement, and promote continuity of care by integrating with public health systems, unlike the segmentation often seen with voluntary EAPs [59]. Limitations of the SROI process include the complexity of assigning financial proxies to sentiment change and the availability of data required for robust calculations, such as displacement and attribution values. A key risk in SROI analysis is an overemphasis on the ratio itself, without considering the underlying content, which provides deeper insight into the value created by different groups [58]. By acknowledging and addressing these challenges through design and analysis strategies, the study generated meaningful insights into the impact of integrating DPS services within an EAP.

Future Research

The study opens several avenues for future research. Long-term impact assessments are needed to evaluate how integrating DPS with EAP services influences therapy utilization and outcomes over extended periods. A broader demographic analysis could provide insights into how factors such as age, socioeconomic status, and cultural background affect engagement with these services. Additionally, incorporating qualitative insights may help illuminate user experiences and satisfaction, offering a deeper understanding of which aspects of the service are most effective or subjectively valued by clients.

Future research could also compare DPS with other digital mental health interventions to identify which models offer the greatest benefits in different contexts. Investigating how DPS integrates with other mental health services, such as primary care or community programs, could provide a more comprehensive understanding of its role within the broader health care system. While comparisons of single- versus multilayered support ecosystems fall beyond the scope of this study, they represent a valuable avenue for future exploration (eg, digital mental health vs in-person services or EAP vs DPS). Additionally, examining DPS's impact on different mental health conditions and conducting a detailed economic analysis of cost savings could further strengthen the case for its adoption. Moreover, this study did not calculate quality-adjusted life years; however, the observed improvements and associated cost reductions provide strong support for the intervention's economic value. Future research could explore additional methods to further quantify these benefits. These efforts would deepen our understanding of DPS' effectiveness and its potential applications across diverse settings.

Conclusions

This study demonstrates that integrating DPS services within an EAP may influence therapy utilization, allowing EAPs to serve higher-acuity clients while lower-acuity clients access supplementary support, such as DPS, for everyday emotional needs. Additionally, the research highlights the potential social, environmental, and broader economic benefits of incorporating DPS services into the emotional support ecosystem, with positive estimated SROI outcomes. By leveraging AI-driven natural language processing for user matching and sentiment analysis, the study underscores the potential of technology to enhance stress management, address life challenges, support mental health, and optimize resource allocation. The findings emphasize the need for further research into DPS effectiveness, particularly regarding long-term impacts, demographic variations, and comparative effectiveness with other interventions. Overall, this study provides valuable insights into the evolving landscape of digital mental health services, reinforcing their role in enhancing organizational wellness programs and identifying key areas for future exploration to maximize their potential.

Conflicts of Interest

Authors HN and ZD disclose an employment relationship at Supportiv. RAM discloses a founder, financial, and an employment relationship at MINES and Associates. Available precautions were taken by all authors to minimize the impact of these affiliations on the study design, data collection, analysis, or interpretation of the findings. All other authors declare no further conflicts of interest.

Multimedia Appendix 1
Breakdown of SROI calculations as calculated using the SROI calculator available on the Sopact website. SROI: social return on investment.
[DOCX File , 8 KB - humanfactors_v12i1e68221_app1.docx]

References

1. World Health Organization (WHO). The world health report. WHO. Geneva, Switzerland: WHO; 2001. URL: https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people [accessed 2025-02-14]

2. Nowrouzi-Kia B, Haritos AM, Long BS, Atikian C, Fiorini LA, Gohar B, et al. Remote work transition amidst COVID-19: impacts on presenteeism, absenteeism, and worker well-being-a scoping review. PLoS One 2024;19(7):e0307087 [FREE Full text] [doi: 10.1371/journal.pone.0307087] [Medline: 39024322]
3. Abramson A. Burnout and stress are everywhere. American Psychological Association. 2022 Jan 1. URL: https://www.apa.org/monitor/2022/01/special-burnout-stress [accessed 2024-10-23]
4. Ling LM. Prevalence of mental health and its impact on employee productivity. Reviews of Contemporary Business Analytics 2023;6(1):1-13 [FREE Full text]
5. Terlizzi E, Schiller J. Mental health treatment among adults aged 18-44: United States, 2019-2021. NCHS Data Brief 2022 Sep(444):1-8 [FREE Full text] [Medline: 36135999]
6. Wu A, Roemer E, Kent K, Ballard D, Goetzel R. Organizational best practices supporting mental health in the workplace. Journal of Occupational & Environmental Medicine 2021;63(12):e925-e931. [doi: 10.1097/jom.0000000000002407]
7. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
8. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff (Millwood) 2013 Feb;32(2):207-214. [doi: 10.1377/hlthaff.2012.1061] [Medline: 23381511]
9. Miele DB, Scholer AA. Self-regulation of motivation. In: Handbook of Motivation at School. New York, NY: Routledge; 2016:375-396.
10. Burns D. The Feeling Good Handbook: The Groundbreaking Program With Powerful New Techniques and Step-by-Step Exercises to Overcome Depression, Conquer Anxiety, and Enjoy Greater Intimacy. Washington, D.C: Plume; 1999.
11. Koechlin H, Coakley R, Schechter N, Werner C, Kossowsky J. The role of emotion regulation in chronic pain: a systematic literature review. J Psychosom Res 2018 Apr;107:38-45. [doi: 10.1016/j.jpsychores.2018.02.002] [Medline: 29502762]
12. Chinman M, George P, Dougherty RH, Daniels AS, Ghose SS, Swift A, et al. Peer support services for individuals with serious mental illnesses: assessing the evidence. Psychiatr Serv 2014 Apr 01;65(4):429-441. [doi: 10.1176/appi.ps.201300244] [Medline: 24549400]
13. Berry LL, Mirabito AM, Baun WB. What's the hard return on employee wellness programs? Harvard Business Review. 2010 Dec 1. URL: https://hbr.org/2010/12/whats-the-hard-return-on-employee-wellness-programs [accessed 2024-10-23]
14. Senn S, Orgera K, Ramirez L. Who's your therapist? Examining the behavioral health workforce. Association of American Medical Colleges. 2023 Oct 26. URL: https://www.aamcresearchinstitute.org/our-work/data-snapshot/whos-your-therapist#:~:text=The%20behavioral%20health%20workforce%20typically,health%20care%20model%20and%20organization [accessed 2025-02-14]
15. Modi H, Orgera K, Grover A. Exploring barriers to mental health care in the U.S. Research and Action Institute. 2022 Oct 10. URL: https://www.aamcresearchinstitute.org/our-work/issue-brief/exploring-barriers-mental-health-care-us [accessed 2024-10-23]
16. Fact sheet: President Biden to announce strategy to address our national mental health crisis, as part of unity agenda in his First State of the Union. The White House. 2022 Mar 1. URL: https://www.whitehouse.gov/briefing-room/statements-releases/2022/03/01/fact-sheet-president-biden-to-announce-strategy-to-address-our-national-mental-health-crisis-as-part-of-unity-agenda-in-his-first-state-of-the-union [accessed 2024-10-23]
17. Tracy K, Wallace S. Benefits of peer support groups in the treatment of addiction. SAR 2016 Sep;Volume 7:143-154. [doi: 10.2147/sar.s81535]
18. Fortuna KL, Venegas M, Umucu E, Mois G, Walker R, Brooks JM. The Future of Peer Support in Digital Psychiatry: Promise, Progress, and Opportunities. Curr Treat Options Psych 2019 Jun 20;6(3):221-231. [doi: 10.1007/s40501-019-00179-7]
19. 24 Million Americans Still Lack Broadband Connectivity. The Pew Charitable Trusts. 2018. URL: https://www.pewtrusts.org/en/research-and-analysis/fact-sheets/2018/07/24-million-americans-still-lack-broadband-connectivity [accessed 2024-10-23]
20. Institute of Medicine. Health literacy: a prescription to end confusion. The National Academies Press. URL: https://nap.nationalacademies.org/catalog/10883/health-literacy-a-prescription-to-end-confusion?onpi_newsdoc04082004= [accessed 2024-10-23]
21. Ryan R, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York, NY: Guilford Publications; 2018.
22. Deci E, Ryan R. Handbook of Self-Determination Research. Rochester, NY: University Rochester Press; 2004.
23. Schwei RJ, Hetzel S, Kim K, Mahoney J, DeYoung K, Frumer J, et al. Peer-to-peer support and changes in health and well-being in older adults over time. JAMA Netw Open 2021 Jun 01;4(6):e2112441. [doi: 10.1001/jamanetworkopen.2021.12441] [Medline: 34129024]
24. Kaplan K, Salzer MS, Solomon P, Brusilovskiy E, Cousounis P. Internet peer support for individuals with psychiatric disabilities: A randomized controlled trial. Social Science & Medicine 2011 Jan;72(1):54-62. [doi: 10.1016/j.socscimed.2010.09.037]

25. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460. [doi: 10.2196/16460] [Medline: 32243256]
26. Rotondi AJ, Haas GL, Anderson CM, Newhill CE, Spring MB, Ganguli R, et al. A clinical trial to test the feasibility of a telehealth psychoeducational intervention for persons with schizophrenia and their families: intervention and 3-month findings. Rehabil Psychol 2005 Nov;50(4):325-336 [FREE Full text] [doi: 10.1037/0090-5550.50.4.325] [Medline: 26321774]
27. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatric Services 2010 Nov 01;61(11):1099-1105. [doi: 10.1176/appi.ps.61.11.1099]
28. Biagianti B, Quraishi SH, Schlosser DA. Potential benefits of incorporating peer-to-peer interactions into digital interventions for psychotic disorders: a systematic review. Psychiatr Serv 2018 Apr 01;69(4):377-388. [doi: 10.1176/appi.ps.201700283] [Medline: 29241435]
29. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Apr;25(2):113-122 [FREE Full text] [doi: 10.1017/S2045796015001067] [Medline: 26744309]
30. Simms L, Ottman KE, Griffith JL, Knight MG, Norris L, Karakcheyeva V, et al. Psychosocial peer support to address mental health and burnout of health care workers affected by COVID-19: a qualitative evaluation. Int J Environ Res Public Health 2023 Mar 03;20(5):4536 [FREE Full text] [doi: 10.3390/ijerph20054536] [Medline: 36901545]
31. Long KH, Moriarty JP, Mittelman MS, Foldes SS. Estimating the potential cost savings from the New York University Caregiver Intervention in Minnesota. Health Aff (Millwood) 2014 Apr;33(4):596-604. [doi: 10.1377/hlthaff.2013.1257] [Medline: 24711320]
32. Bassuk EL, Hanson J, Greene RN, Richard M, Laudet A. Peer-delivered recovery support services for addictions in the United States: a systematic review. J Subst Abuse Treat 2016 Apr;63:1-9. [doi: 10.1016/j.jsat.2016.01.003] [Medline: 26882891]
33. Rayland A, Andrews J. From Social Network to Peer Support Network: Opportunities to Explore Mechanisms of Online Peer Support for Mental Health. JMIR Ment Health 2023 Feb 28;10:e41855. [doi: 10.2196/41855]
34. Madaan A, Zhou S, Alon U, Yang Y, Neubig G. Language models of code are few-shot commonsense learners. In: Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. New York, NY: Association for Computing Machinery; 2022 Presented at: The 2022 Conference on Empirical Methods in Natural Language Processing; December 7-11, 2022; Abu Dhabi, United Arab Emirates. [doi: 10.18653/v1/2022.emnlp-main.90]
35. Engel G. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 08;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]
36. Nicholls J, Lawlor E, Neitzert E, Goodspeed T. A guide to social return on investment (SROI) (Revised). Socioeco.org. 2012. URL: https://www.socioeco.org/bdf_fiche-publication-929_en.html [accessed 2024-10-23]
37. Millar R, Hall K. Social Return on Investment (SROI) and Performance Measurement. Public Management Review 2013 Sep;15(6):923-941. [doi: 10.1080/14719037.2012.698857]
38. Engel GL. The clinical application of the biopsychosocial model. Am J Psychiatry 1980 May;137(5):535-544. [doi: 10.1176/ajp.137.5.535] [Medline: 7369396]
39. Elyoseph Z, Refoua E, Asraf K, Lvovsky M, Shimoni Y, Hadar-Shoval D. Capacity of Generative AI to Interpret Human Emotions From Visual and Textual Data: Pilot Evaluation Study. JMIR Ment Health 2024 Feb 06;11:e54369 [FREE Full text] [doi: 10.2196/54369] [Medline: 38319707]
40. Dana Z, Nagra H, Kilby K. Role of synchronous, moderated, and anonymous peer support chats on reducing momentary loneliness in older adults: retrospective observational study. JMIR Form Res 2024 Oct 25;8:e59501 [FREE Full text] [doi: 10.2196/59501] [Medline: 39453688]
41. Social return on investment (SROI) - new approach. Sopact. URL: https://www.sopact.com/guides/social-return-on-invesment-sroi#social-return-on-investment-sroi [accessed 2024-10-23]
42. Greene J, Hibbard JH. Why does patient activation matter? An examination of the relationships between patient activation and health-related outcomes. J Gen Intern Med 2012 May;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
43. Schunk DH, DiBenedetto MK. Motivation and social cognitive theory. Contemporary Educational Psychology 2020 Jan;60:101832. [doi: 10.1016/j.cedpsych.2019.101832]
44. Lin M, Weng W, Apriliyasari R, VAN Truong P, Tsai P. Effects of patient activation intervention on chronic diseases: a meta-analysis. J Nurs Res 2020 Oct;28(5):e116. [doi: 10.1097/jnr.0000000000000387] [Medline: 32649394]
45. Taylor V. Measuring Healthy Days: Population Assessment of Health-Related Quality of Life. PsycEXTRA Dataset 2000:40 [FREE Full text]
46. Revicki DA, Gnanasakthy A, Weinfurt K. Documenting the rationale and psychometric characteristics of patient reported outcomes for labeling and promotional claims: the PRO Evidence Dossier. Qual Life Res 2007 May 1;16(4):717-723. [doi: 10.1007/s11136-006-9153-5] [Medline: 17268927]

47. Golden-Biddle K. A Review of: "Positive Organizational Scholarship: Foundations of a New Discipline, edited by Kim S. Cameron, Jane E. Dutton, and Robert E. Quinn". International Public Management Journal 2007 Feb 24;9(2):199-202. [doi: 10.1080/10967490600767035]
48. Ilies R, Dimotakis N, De Pater I. Psychological and physiological reactions to high workloads: implications for well - being. Personnel Psychology 2010 May 12;63(2):407-436. [doi: 10.1111/j.1744-6570.2010.01175.x]
49. Cross S, Bell I, Nicholas J, Valentine L, Mangelsdorf S, Baker S, et al. Use of AI in Mental Health Care: Community and Mental Health Professionals Survey. JMIR Ment Health 2024 Oct 11;11:e60589-e60589. [doi: 10.2196/60589]
50. Bossewitch J, Brown LX, Gooding PM, Harris L, Horton J, Katterl S, et al. SSRN Journal 2022:4215994 [FREE Full text] [doi: 10.2139/ssrn.4215994]
51. Stigma, prejudice and discrimination against people with mental illness. Psychiatry. URL: https://www.psychiatry.org/patients-families/stigma-and-discrimination [accessed 2024-10-23]
52. Duckett LJ. Quantitative Research Excellence: Study Design and Reliable and Valid Measurement of Variables. J Hum Lact 2021 Aug 09;37(3):456-463. [doi: 10.1177/08903344211019285] [Medline: 34107788]
53. Mara C, Peugh J. Validity of data collected from randomized behavioral clinical trials during the COVID-19 pandemic. J Pediatr Psychol 2020 Oct 01;45(9):971-976. [doi: 10.1093/jpepsy/jsaa078] [Medline: 32968774]
54. Harder VS, Stuart EA, Anthony JC. Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. Psychol Methods 2010 Sep;15(3):234-249 [FREE Full text] [doi: 10.1037/a0019623] [Medline: 20822250]
55. ijones. Measurement error and bias. In: Epidemiology for the Uninitiated. London, UK: BMJ; 2020.
56. Marr B. Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems. Hoboken, NJ: John Wiley & Sons; 2019.
57. Thieme A, Hanratty M, Lyons M, Palacios J, Marques RF, Morrison C, et al. Designing human-centered AI for mental health: developing clinically relevant applications for online CBT treatment. ACM Trans Comput-Hum Interact 2023 Mar 17;30(2):1-50. [doi: 10.1145/3564752]
58. Attridge M. A global perspective on promoting workplace mental health and the role of employee assistance programs. Am J Health Promot 2019 May 21;33(4):622-629. [doi: 10.1177/0890117119838101c] [Medline: 31006254]
59. Nordic cooperation. URL: https://www.norden.org/en/nordic-council-ministers [accessed 2025-01-05]

Abbreviations

AI: artificial intelligence
DPS: digital peer support
EAP: employee assistance program
HRQoL: health-related quality of life
PSM: peer support moderator
ROI: return on investment
SF-36: 36-item Short Form
SROI: social return on investment

Edited by A Kushniruk; submitted 31.10.24; peer-reviewed by I Pyykko, CI Sartorão Filho, KW Tay; comments to author 22.12.24; revised version received 09.01.25; accepted 28.01.25; published 25.02.25.

Please cite as:
Nagra H, Mines RA, Dana Z
Exploring the Impact of Digital Peer Support Services on Meeting Unmet Needs Within an Employee Assistance Program: Retrospective Cohort Study
JMIR Hum Factors 2025;12:e68221
URL: https://humanfactors.jmir.org/2025/1/e68221
doi:10.2196/68221
PMID:39998863

Original Paper

Identifying Strategies for Home Management of Ostomy Care: Content Analysis of YouTube

Marketa Haughey[1*], BSc, PhD; David M Neyens[2*], PhD; Casey S Hopkins[3*], PhD; Christofer Gonzaga[2*], BS; Melinda Harman[1*], PhD

[1]Department of Bioengineering, Clemson University, Clemson, SC, United States

[2]Department of Industrial Engineering, Clemson University, Clemson, SC, United States

[3]School of Nursing, Clemson University, Clemson, SC, United States

[*]all authors contributed equally

Corresponding Author:
Melinda Harman, PhD
Department of Bioengineering
Clemson University
301 Rhodes Engineering Research Center
Clemson, SC, 29634
United States
Phone: 1 8646564140
Email: harman2@clemson.edu

Abstract

Background: The social media platform YouTube is a recognized educational resource for health information, but few studies have explored its value for conveying the lived experience of individuals managing chronic health conditions and end users' interactions with medical device technology. Our study explores self-care strategies and end user needs of people living with a stoma because patient education and engagement in ostomy self-care are essential for avoiding ostomy-related complications. Ostomy surgery creates a stoma (an opening) in the abdomen to alter the route of excreta from digestive and urinary organs into a detachable external pouching system. After hospital discharge, people who have undergone ostomies perform critical self-care tasks including frequent ostomy appliance changes and stomal and peristomal skin maintenance.

Objective: The purpose of this study was to systematically assess YouTube videos narrated by people who have undergone ostomies about their ostomy self-care in home (nonhospital) settings with a focus on identifying end user needs and different strategies used by people who have undergone ostomies during critical self-care tasks.

Methods: Using predefined search terms and clear inclusion and exclusion criteria, we identified YouTube videos depicting narrators who have undergone ostomies and their ostomy self-care in home settings. Using a consensus coding approach among 3 independent reviewers, all videos were analyzed to collect metadata, data of narrators who have undergone ostomies, and specific content data.

Results: There were 65 user-generated YouTube videos that met the inclusion and exclusion criteria. These videos were posted by 28 unique content creators representing a broad range of ages who used a variety of supplies. The common challenges discussed were peristomal skin complications, inadequate appliance adhesion and subsequent leakage, and supplies-related challenges. Narrators who have undergone ostomies discussed various expert tricks and tips to successfully combat these challenges.

Conclusions: This study used a novel approach to gain insights about end user interactions with medical devices while performing ostomy self-care, which are difficult to gain using traditional behavioral techniques. The analysis revealed that people who have undergone ostomies are willing to share their personal experience with ostomy self-care on the web and that these videos are viewed by the public. User-generated videos demonstrated a variety of supplies used, end user needs, and different strategies for performing ostomy self-care. Future research should examine how these findings connect to YouTube ostomy self-care content generated by health care professionals and organizations and to guidelines for ostomy self-care.

(JMIR Hum Factors 2025;12:e66634) doi:10.2196/66634

KEYWORDS

medical device usability; digital health; online support groups; living with chronic medical conditions; ostomy self-care; YouTube; patient education; user needs assessment; users experience; social media; ostomates; colostomy; ileostomy; usability; usefulness; utility; wearable device; medical device; support group; socials; social network; ostomy; digital; digital technology; digital intervention

Introduction

YouTube as Health Information Resource

Individuals are increasingly seeking health-related information on the web, and in some cases, patients rely on the internet as much or more than their physicians [1,2]. Patients have also reported using web-based resources to make health-related decisions and to manage chronic conditions [2,3].

YouTube, the largest on the web video-sharing platform worldwide for streaming a variety of user-generated content with 2.5 billion users worldwide [4] is a recognized educational resource for sharing and disseminating health-related information and influencing individuals' viewpoints related to health care topics such as disease prevention, treatment therapy, or immunizations [2,3]. Several studies highlight the informational flaws and biases in a wide variety of YouTube health information videos related to clinical decision-making, diagnosing, treatment recommendation, and health promotion. The analysis in these studies generally focuses on the health information provided rather than the individual user experience with the health issue [2,5-9].

Studies involving so-called "e-patients" who are motivated to share health-related personal experiences on the web reveal that sharing their lived experience and knowledge of their disease can empower and engage others while also providing (and gaining) social support [10,11]. While many health-related government organizations such as the World Health Organization, the National Institute of Health, the Center for Disease Control and Prevention, and the American Red Cross use YouTube to disseminate health information [1], user-generated videos are also used as an educational resource. This suggests that user-generated content decreases the gap between health information delivered by health care professionals and the lived experience of dealing with a disease.

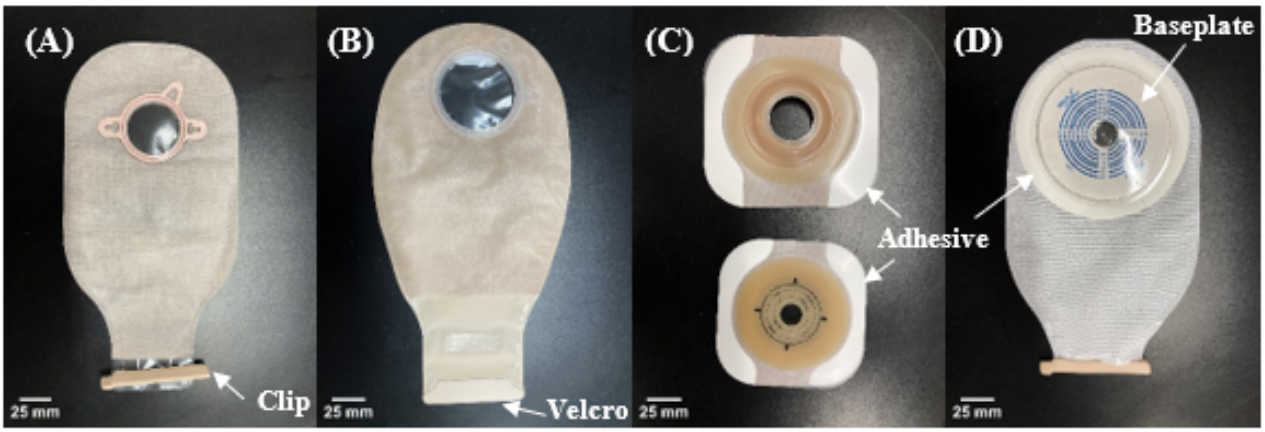
Medical Device Usability in the Home Setting

Usability studies examining the interaction between medical device technology and end users are essential for developing medical devices that are safe and effective [12]. These studies allow for actively including end users in the design process while considering end user competencies and environment, thus, continuously improving medical device design to fit user needs [13]. Capturing the real world, end user experience of people with chronic conditions and their interaction with complex medical device technology in home (nonhospital) settings is challenging [12], such as having a stoma and routinely using an ostomy appliance, and to our knowledge, these studies are limited [14].

Ostomy Surgery

There are about one million people living with ostomy in the United States and more than 150,000 Americans undergo ostomy surgery each year [15,16]. Ostomy surgery alters the route of excreta from internal digestive and urinary organs through a stoma in the abdomen and into a detachable external pouching system (Figure 1) that fits snugly around the stoma [16]. People who have undergone ostomies perform critical self-care tasks that include frequent changing of the pouching system (3-10 changes per week) and maintaining healthy stoma and peristomal skin [17]. Therefore, there is a need to understand the usability of medical device technology used by people who have undergone ostomies in ostomy appliance change procedures and self-care.

Figure 1. An ostomy appliance consists of an external pouching system that is attached to the abdomen of people who have undergone an ostomy using adhesive. There are various design features, including a (A) 2-piece pouch with clip closure, (B) 2-piece pouch with Velcro closure, (C) precut (top) and moldable (bottom) baseplate that is subsequently attached to the 2-piece pouch [in (A) or (B)] at the ring, and (D) 1-piece appliance with clip closure (pouch and baseplate being 1 part).



Ostomy Care Patient Education and Self-Management

It is widely recognized that patient education and engagement in ostomy self-care are essential for avoiding ostomy-related complications [18-20]. However, the education of new people who have undergone ostomies to gain the knowledge and skills needed for managing chronic conditions and performing self-care tasks is not well defined [18]. Ideally, according to the Wound, Ostomy, and Continence Nurses Society Clinical Guidelines and recommendations [21], new people who have undergone ostomies should obtain education from ostomy nurse specialists. Nevertheless, there is evidence of insufficient pre- and postoperative education provided to new people who have undergone ostomies related to stoma and peristomal skin care and complications [22-25]. Due to the short hospitalization trends and the limited availability of ostomy nurse specialists, there is limited teaching time and formal follow-up with the specialists [22-25]. Further, new people who have undergone ostomies may not be mentally and physically prepared for self-management of their ostomy, which requires gaining new knowledge, skills, and attitudes [26,27]. Most people who have undergone ostomies experience some long-term challenges concerning the management of daily self-care of their stoma that broadly impact their health-related quality of life and outcomes [28,29]. Peristomal skin complications are a common problem following ostomy surgery affecting over one-third of patients with ostomies within 90 days post surgery and up to 80% within 2 years [30-33]. While it is known that skin problems interfere with pouch adhesion, causing challenges with pouch leakage, and thus odor, and the necessity for frequent and unscheduled appliance changes [29], it is very difficult to study real-world responses to such challenges.

Study Objective

Given the complexity of ostomy appliance change procedure medical technology and self-care, there is a need to assess the usability of this technology and how self-management skills are developed by people who have undergone ostomies over time. We completed a systematic analysis of user-generated videos related to ostomy self-care published on YouTube and created by individuals with expertise in managing chronic conditions about ostomy self-care in home (eg, nonhospital) settings. Our objective was to identify end user experiences with ostomy medical device technology and the different strategies and procedures used by these experts to change their ostomy appliance and perform stomal and peristomal skin care.

Methods

Ethical Considerations

This study was excluded from an institutional ethics board review because it assessed only copyrighted, public source information and no individual user data was included in the manuscript.

Search Terms and Search Strategy

Search terms related to ostomy self-care were defined by the research team (Table 1) for identifying videos using the search tools within YouTube. During December 2023, YouTube was searched for relevant videos using search terms and Boolean operators under the search category “videos only” tab or “video” filter. All searches were conducted using an incognito or private browser that did not retain cookies to prevent suggested related videos from appearing on search results and to ensure that the search for each term could be replicated. The first 20 videos for each search term were reviewed and documented in a spreadsheet. No videos were included in the analysis if they appeared in the YouTube search result under the “people also watched,” “shorts,” or “sponsored websites” section as these videos may or may not have met the search term criteria. The initial search resulted in 960 videos.

Table 1. Search terms used to search for relevant user-generated videos published on YouTube.

Search term category	Search terms
Ostomy pouch change procedure	[a] + (pouch OR bag) + (care OR change OR procedure)
Ostomy tips and tricks	[a] + (pouch OR bag) + (pros and cons OR tips OR tips and tricks OR empty OR cleaning OR care OR leakage OR leakage prevention)
Ostomy self-care and home management	[a] + (pouch change OR bag change) + (home OR self-care OR care)
Ostomy skincare	[a] + (skin OR peristomal skin) + (care OR treatment OR irritation OR dermatitis)

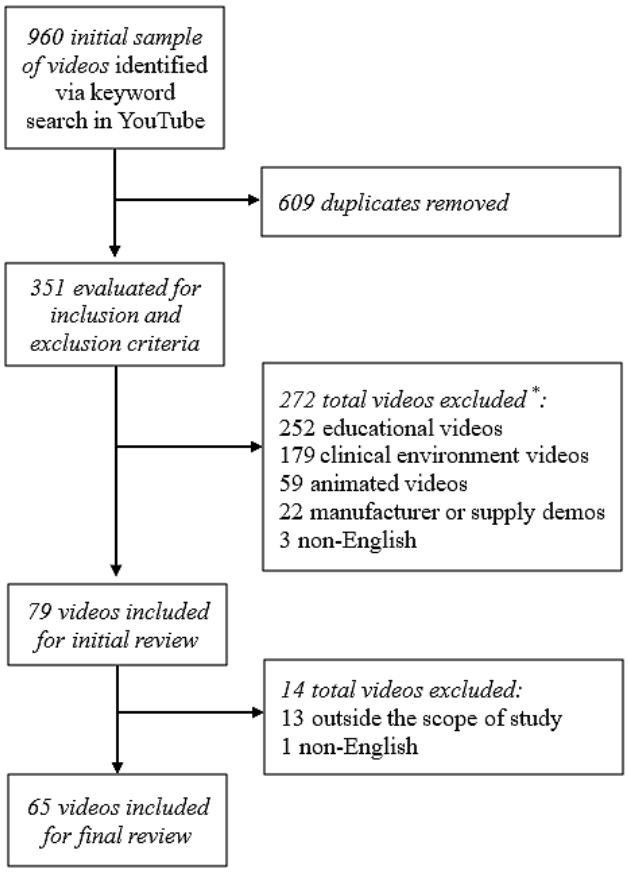
[a]Is ostomy OR urostomy OR colostomy OR ileostomy OR stoma OR colorectal cancer OR Crohn disease OR inflammatory bowel disease OR IBD OR diverticulitis OR bladder cancer OR cystectomy.

Inclusion and Exclusion Criteria

Overview

From the initial set of 960 videos, a final set of 65 videos (Figure 2) was identified for review after the removal of duplicates and the application of inclusion and exclusion criteria (Textbox 1). Duplicate videos (n=609) were identified and removed by channel grouping and comparing URLs and video titles. The inclusion and exclusion criteria were applied to eliminate educational videos (n=252), videos recorded in a clinical environment (n=179), animated videos (n=59), supply or manufacturer demonstration videos (n=22), videos in languages other than English (n=4) and those videos that were outside the scope of this study (n=13) such as broad discussions of clinical issues or ostomy support. Educational videos were defined as videos not discussing personal experiences related to ostomy self-care, for example, videos providing information related to ostomy belts.

Figure 2. Video search inclusion and exclusion process. *Some videos were excluded for multiple criteria.



Textbox 1. Specific inclusion and exclusion criteria.

Included videos

- Were found using search terms
- Included a user narrative
- Involved any channel, age, gender, race, ethnicity, year, or length
- Depicted ostomy appliance change procedures in a nonhospital setting
- Discussed ostomy self-care, tips, tricks, or challenges
- Involved a person living with a stoma as a narrator
- Were published in the English language

Excluded videos

- Were recorded in a clinical environment or were narrated by a medical professional (not having a stoma)
- Included experiences of users who have not undergone ostomies
- Were educational videos
- Were supply or manufacturer demonstration videos
- Were animated
- Duplicated videos from other search terms
- Were published in the non-English language
- Categorized as video “shorts”
- Were outside the scope of this study

Data Coding

In total, 3 categories of video-specific variables were recorded, including video metadata, narrator data, and specific content data (Table 2). The researchers used consensus coding of the included videos, which is a well-accepted method to enhance credibility and reliability in qualitative research [34]. Before video reviews, 3 general categories and 15 variables were named and defined (Table 2). Each video was then coded individually by 2 reviewers using the defined categories and any discrepancies between reviewers were discussed. If consensus was not reached, a third reviewer reviewed the video and data coding to address bias and reach consensus and to ensure accuracy in data coding. The category definitions were then revised as needed before a final consensus coding of all the included videos by 2 reviewers.

Table 2. Data coding for 3 categories of videos including specific variables, and their definition and type.

Category and variable	Definition
Video metadata	
Title	The title of the video by the content creator
URL	The web address of the video
Channel	The name of the page where videos are posted for viewing
Posted year	The year that video was publicly posted to the YouTube platform
Video length	The time duration of the video
View counts	The number of times the video was viewed as of January 4, 2024
Public ratings	The number of likes or dislikes as of January 4, 2024
Data of narrators who have undergone ostomies	
Presenting assumed gender	The assumed gender of the content creator based on gender markers such as physical build, voice, clothes, and hair
Presenting assumed age group	The assumed age group of the content creator mainly based on physical appearance and hair coloration
Specific video content data	
Supplies used	Any item, product, equipment, or accessory used or discussed as used by a narrator who has undergone an ostomy during the ostomy appliance change procedure
Appliance change frequency[a]	How frequently the narrator who has undergone an ostomy completed an appliance change process
Type of ostomy[a,b]	The type of ostomy disclosed by the narrator who has undergone an ostomy
Underlying condition for ostomy[a,b]	The medical condition experienced by the narrator who has undergone an ostomy and contributing to the ostomy surgery
Common challenges[a]	Description of the problems related to ostomy, ostomy appliance change procedures, and ostomy self-care
Expert tips and tricks	Advice or recommendation given by a narrator who has undergone an ostomy to the audience to address common challenges, aid in expediting or easing the appliance change process, or provide additional information to improve the overall user experience

[a]Variable was not communicated in every video.

[b]If not communicated in the video, the intro page and the about me page for the creator were reviewed.

Results

Video Metadata

After removing duplicate videos and applying the inclusion and exclusion criteria, there were 65 videos that explored different topics related to ostomy, ostomy appliance change procedures, and the self-care of people who have undergone ostomies. Included videos were posted between 2011 and 2023 and averaged about 9.5 minutes in length, over a quarter-million views, and more than 3000 likes (Table 3).

Table 3. The video metadata descriptive statistics for the videos included in the analysis.

Video metadata	Data
Posted year, n/N (%)	Between 2011 and 2023 (11/65, 17% and 12/65, 18% videos in 2020 and 2023, respectively)
Video length, mean (SD; range)	9:36 min (SD 6:30 s; 1:00-31:40 min)
View counts, mean (SD; range)	269,076 (SD 922,352; 1900-5,493,702) views
Public ratings, mean (SD; range)	3015 (SD 8761; 0-58,000) likes

Data of Narrators Who Have Undergone Ostomies

There were 28 unique content creators with several individual creators posting multiple videos. Further, 2 high volume narrators who have undergone ostomies uploaded almost half (31/65) of all the videos (24 and 7 videos from the 2 most frequent content creators). Most videos included a female narrator and a narrator in the 20-30 years age range followed by a middle-aged narrator, an older narrator, and a teenage narrator (Table 4). None of the narrators who have undergone ostomies communicated their gender or age in the videos. Therefore, the presenting assumed gender was determined by the reviewers based on the gender markers such as physical build, voice, clothes, and hair and the presenting assumed age based on the physical appearance and hair coloration.

Table 4. Data of narrators who have undergone ostomies for videos included in the final review.

Unique narrators who have undergone ostomies (n=28)	Data, n/N (%)
Presenting assumed gender	
Female	20/28 (71% narrators who have undergone ostomies)
Male	8/28 (29% narrators who have undergone ostomies)
Presenting assumed age group	
Infant[a]	1/65 (2% videos)
Teenager	4/65 (6% videos)
Aged 20-30 years	42/65 (65% videos)
Middle age	13/65 (20% videos)
Older	5/65 (8% videos)

[a]One video was presented about an infant with an ostomy that was narrated by the parent caregiver.

Specific Video Content Data

Ostomy Video Content

From the final set of 65 videos, 26/65 (40%) videos discussed ostomy self-care tips and tricks that focused most frequently on maintaining peristomal skin health, preventing leaks, pancaking, supply use, and general tips; 21/65 (32%) videos captured a full ostomy appliance change procedure and 14/65 (22%) videos included part of an ostomy appliance change procedure such as emptying, specific supply use, and skin preparation and appliance application, among others; and 4/65 (6%) videos were about the personal experience of a person living with a stoma (Table 5).

Table 5. Types of ostomy-related video content, its frequency, and topics discussed.

Video content and topic discussed	Video count, n	Video frequency, n/N (%)
Tips and tricks		26/65 (40%)
Peristomal skin health	6	
Leaks	5	
Pancaking	3	
Emptying or not emptying	3	
Specific supply use	2	
General	2	
Odor	1	
Intimacy	1	
Adhesion	1	
Pouch decoration	1	
Full ostomy appliance change		
Ostomy appliance change procedure	21	21/65 (32%)
Part of ostomy appliance change		14/65 (22%)
Empty an appliance	5	
Specific supply use	3	
Skin preparation and appliance application	2	
Change appliance after shower	1	
Remove appliance and clean peristomal skin	1	
Cut or fit ostomy baseplate	1	
Appliance change omitting details	1	
Experience		4/65 (6%)
Living with stoma	2	
Living in London	1	
Stories of people who have undergone ostomies	1	

Supplies Used

The narrators who have undergone ostomies did not all use identical supplies and there was a variety of types, numbers, and frequencies of supplies used during the ostomy appliance change procedures (Table 6). Among the 35 videos that captured a full or partial ostomy appliance change procedure, 25/35 (71%) included a 2-piece pouching system, and 9/35 (26%) included a 1-piece pouching system (Table 6). In 1/35 (3%) videos, the type of ostomy appliance could not be determined. Furthermore, the baseplate was cut to fit or precut by the supplier in 25/35 (71%) videos and molded by hand in 3/35 (9%) videos. In 7/35 (20%) videos, the method for sizing the baseplate by cutting or molding could not be determined.

Further, supplies used to clean the peristomal skin and remove any remaining adhesive or output included paper towel or toilet paper in 19/35 (54%) videos, adhesive remover wipes in 17/35 (49%) videos, adhesive remover spray in 14/35 (40%) videos, dry wipes in 11/35 (31%) videos, towel or washcloth in 9/35 (26%) videos, water in 7/35 (20%) videos, and wet wipes in 4/35 (11%) videos. Supplies used to protect the skin included a barrier ring in 17/35 (49%) videos, skin barrier in the form of wipe, spray, or cream in 15/35 (43%) videos, and ostomy paste in 6/35 (17%) videos. Supplies used to treat peristomal skin disorders included stoma powder in 12/35 (34%) videos and skin treatments (eg, corticosteroids, and antifungal or other anti-inflammatory medications) in 8/35 (23%) videos. Supplies used to aid adhesion between the pouch and peristomal skin included extenders in 5/35 (14%) videos and belts in 4/35 (11%) videos. Supplies used to prevent or minimize odor issues were odor eliminators in 9/35 (26%) videos and lubricating deodorant in 5/35 (14%) videos. Other general ostomy supplies that were used included plastic bags in 19/35 (54%) to discard the ostomy appliance before disposal, scissors in 16/35 (46%) videos to cut a baseplate to match the size of the stoma, some kind of bag in 4/35 (11%) videos to carry ostomy supplies, and some type of container in 4/35 (11%) videos to rinse the pouch with water.

Table 6. Supplies used or discussed in the 35 videos depicting full or partial ostomy appliance change procedures organized by supply category.

Supply category and supply	Videos[a], n (%)
Pouching system	
2-piece appliance	25 (71)
1-piece appliance	9 (26)
Not reported	1 (3)
Stoma or peristomal skin cleaning	
Paper towel or toilet paper	19 (54)
Adhesive remover wipes	17 (49)
Adhesive remover spray	14 (40)
Dry wipes	11 (31)
Towel or washcloth	9 (26)
Water	7 (20)
Wet wipes	4 (11)
Gauze	2 (6)
Skin cleanser	2 (6)
Soap	2 (6)
Saline	1 (3)
Peristomal skin protection	
Barrier ring	17 (49)
Skin barrier (wipe, spray, or cream)	15 (43)
Ostomy paste	6 (17)
Flow assist device	2 (6)
Peristomal skin treatment	
Stoma powder	12 (34)
Skin treatments	8 (23)
Pouch adhesion support	
Extenders	5 (14)
Belt	4 (11)
Skin adhesive	3 (9)
Equalizer	2 (6)
Heating pad	2 (6)
Odor elimination	
Odor eliminator (deodorant or air freshener)	9 (26)
Lubricating deodorant	5 (14)
Lubricant	2 (6)
General	
Plastic bags	19 (54)
Scissors	16 (46)
Supply bag or bag or ziplock	4 (11)
Plastic container or water bottle or cup	4 (11)
Marker	3 (9)
Size template	3 (9)
Disposable pad	1 (3)

Supply category and supply	Videos[a], n (%)
Q-tip	1 (3)
Shower stopper	1 (3)
Syringe	1 (3)

[a]Calculated as x/35×100%.

Underlying Condition and Type of Ostomy

One-third of the narrators who have undergone ostomies (9/28, 32%) did not disclose their underlying condition. Crohn disease was the most frequently reported underlying condition for 10/28 (36%) narrators who have undergone ostomies followed by ulcerative colitis (3/28, 11%), cancer (3/28, 11%), and unspecified inflammatory bowel disease (1/28). Further, 2 (7%) narrators who have undergone ostomies reported other underlying conditions for their ostomy surgery. Per the type of ostomy, ileostomies accounted for 79% (22/28) and colostomies accounted for 14% (4/28). Furthermore, 2 narrators who have undergone ostomies did not report their type of ostomy and no urostomy videos met inclusion and exclusion criteria for final review.

Appliance Change Frequency

More than half of narrators who have undergone ostomies (16/28, 57%) did not discuss appliance change frequency. For the remaining narrators who have undergone ostomies (12/28, 43%), there was considerable variability in ostomy appliance change frequency. In general, a majority of narrators who have undergone ostomies (n=10) reported changing their appliance within a 2-5-day interval, 2 every 7 and 15 days, and 1 when the pouch is full. Further, 1 narrator reported various changing frequencies across 4 videos (2-3 days, 3-4 days, every 4 days, and 4-5 days). Additionally, 2 narrators who have undergone ostomies commented on changing the appliance more frequently when having peristomal skin issues (every day or every other day) and emptying the ostomy appliance 5 to 8 times a day.

Common Challenges

About one-third of the content creators noted specific issues with supplies or processes that required workarounds or techniques to be implemented. Major challenges included peristomal skin complications (21/65 videos, 32%), leaks around the adhesive area (20/65 videos, 31%), inadequate appliance adhesion (13/65 videos, 20%), and no control over stoma effluent discharge making it difficult to keep the stoma and peristomal skin clean while changing the ostomy appliance (13/65 videos, 20%). Within the videos, the narrators who have undergone ostomies reported peristomal skin complications related to skin irritation and redness due to the output leakage, peristomal skin disorders such as granulomas, or allergic reactions toward specific ostomy supplies and these complications subsequently compromised appliance adhesion, caused pain, and led to further leaks.

Narrators who have undergone ostomies also expressed difficulties related to supplies in 17/65 (26%) videos. For example, supplies leaving residues on the skin compromising appliance adhesion, supplies containing alcohol causing a burning sensation on the wounded skin, variability in supplies' names causing confusion, and supplies not delivered on time or not working properly for a given individual. Challenges with the design of the ostomy appliance were noted, including the clip of the pouch being uncomfortable, difficulty with cleaning the end of the pouch after emptying, moisture absorption by the barrier ring leading to inadequate adhesion, or a 2-piece appliance being more difficult to apply. Finally, pouch pancaking with the output being trapped at the top of the pouch, pouch ballooning due to the gas accumulating in the ostomy appliance, cutting the baseplate to fit the stoma size, emptying an ostomy appliance and its cleaning, lifestyle adjustments (clothing, diet, relationships, intimacy, travelling, etc), and odor were additional challenges discussed in the videos.

Expert Tips and Tricks

The narrators who have undergone ostomies identified various strategies to address the challenges encountered. To overcome peristomal skin complications, narrators who have undergone ostomies (1) emphasized the importance of cleaning peristomal skin (eg, shower without an appliance and using adhesive remover to remove any output residue which could further irritate the skin); (2) avoided supplies containing alcohol to minimize burning sensations on the skin; (3) used a crusting technique which involves the application of stoma powder over sore skin followed by moistening the powder with skin barrier (wipe or spray), letting the powder dry, and repeating the process multiple times [16]; (4) used ostomy paste, or barrier ring to protect and allow healing of the wounded skin by creating a protective barrier between the wounded skin and appliance, or implemented more aggressive skin treatment products such as steroids or silver nitrate based on the consultation with ostomy care nurse specialists; and (5) recommended balance between changing an appliance too frequently or too rarely.

The tips of narrators who have undergone ostomies to avoid leakage included (1) properly fitting the appliance baseplate with the sizing of the opening for the stoma not to be too tight or too big; (2) using a barrier ring or ostomy paste to fill any abdominal gaps and divots; (3) using an ostomy belt, extenders, or medical tape to support the weight of the pouch; (4) trying different supplies (eg, convex pouches for retracted stomas) and different suppliers to identify supplies working properly for a given individual, while also emphasizing not using too many supplies; and (5) emptying an appliance regularly.

Narrators who have undergone ostomies believed that appliance adhesion was improved by (1) cleaning the skin properly using an adhesive remover followed by cleaning with water or some type of wipe to avoid leaving any residues that could compromise adhesion, (2) ensuring the peristomal skin is dry before applying an appliance, (3) wearing an ostomy belt, (4)

using an equalizer ring that applies an equal pressure around the stoma or ostomy paste that fills any divots or gaps between appliance and abdomen, (5) applying an appliance on a flat abdominal surface (eg, leaning backward or standing), (6) massaging the baseplate and running a finger around the stoma, and (7) using heat to stimulate baseplate adhesion to the abdomen (eg, hairdryer on low setting, heating pad, or holding hand over pouch). Narrators who have undergone ostomies also acknowledged there is not only one proper way of performing ostomy self-care and that what works for them may not work for someone else. Thus, they recommended trying different supplies and self-care strategies and consulting with ostomy nursing care specialists about supplies selection and their use for skin care.

Concerning managing a frequent and uncontrollable stoma output, some narrators who have undergone ostomies found it effective to eat marshmallows or to fast before changing an appliance to slow down the output, covering the stoma with some type of wipe, paper towel, or gauze while changing the appliance to protect the skin and make it easier to clean, and organizing or preparing all supplies and emptying the pouch before starting to change the appliance.

Narrators who have undergone ostomies conveyed strategies to prevent "pancaking" such as using a lubricant or some type of oil inside the pouch, emptying and rinsing the pouch with water, disabling an appliance filter, and blowing air into the pouch. Further, they used deodorants and "burped" (let the air out of the pouch) their appliance to prevent ballooning. They used an ostomy cover and deodorizing drops along with eating naturally deodorizing food such as kefir, yoghurt, or parsley to deal with the odor. To overcome the challenge of cutting a baseplate to the proper size, narrators who have undergone ostomies traced the stoma size with a marker on the template and saved the template with a marking for repeated use, or ordered the appliance precut from suppliers once their stoma size remained constant. Strategies used to adjust to the postostomy lifestyle included (1) finding and joining a support group, (2) talking to people who "get it," (3) learning about their stomas (eg, diversity of supplies, proper diet and hydration, or clothing tips), (4) exercising to reduce anxiety, and (5) developing an ability to perform ostomy self-care while also having someone else who can change their appliance.

Discussion

Principal Findings

This study uses a novel approach for gaining insight into ostomy care end user (person who has undergone an ostomy) experience and usability with medical device (ostomy appliance change) technology through the identification of strategies and supplies used, and common challenges and needs experienced by people who have undergone ostomies while changing the ostomy appliance and performing self-care. The needs, supply choices, and self-care strategies that people who have undergone ostomies used and discussed reflected the need to address a common challenge of peristomal skin complications following an ostomy surgery [21-23] and demonstrated characteristics of an expert understanding [24-27].

YouTube as Ostomy Self-Care Resource

This study examined YouTube videos related to the interaction of people who have undergone ostomies with medical device technology and self-care that were provided by and for the public, and the need for additional information related to long-term ostomy self-care in home (nonhospital) settings. In our study, we initially identified 960 ostomy-related videos and analyzed 65 videos with a narrator who has undergone an ostomy that met our inclusion criteria. Each of these videos was viewed 269,000 times on average, liked 3000 times on average, and posted by 28 unique content creators representing a broad range of ages and both, female and male genders. Our study supports current research suggesting that individuals are increasingly seeking health-related information on the web and patients use web-based resources to make health-related decisions and to manage chronic conditions [1-3,35,36]. Patient education and engagement in ostomy self-care are essential for performing self-care tasks and successful management of their chronic condition [18-20,35]. Nevertheless, research evidence shows the pre- and postoperative education provided to people who have undergone ostomies is insufficient [22-25] and new people who have undergone ostomies may not be prepared to gain new knowledge, skills, and attitudes necessary for self-care management of their ostomy immediately post surgery [26,27]. Therefore, our study suggests patients are looking for additional support and resources for ostomy self-care.

The video metadata and data of narrators who have undergone ostomies show that people who have undergone ostomies of various ages and genders are willing to share their personal experiences with ostomy self-care and appliance change procedures on the web and that these videos are viewed by others. Through YouTube, people who have undergone ostomies can share their stories, empower and help other people who have undergone ostomies, and develop a sense of community with other people who have undergone ostomies [10,11], which goes beyond the ability of health care providers not living with a stoma. In our study, narrators who have undergone ostomies created videos depicting ostomy self-care not to substitute health care professional guidelines and recommended practices, but rather to share a lived experience with a chronic illness, and their challenges and wins. They also commented on being able to develop a sense of community, learn from the lived experiences of other people who have undergone ostomies, and adjust to postostomy surgery lifestyle changes such as by finding and joining a support group or talking to people who "get it." Many of the narrators who have undergone ostomies made disclosures that they were not giving medical advice and commented on positive experiences with and encouraged interaction with ostomy care nurse specialists. The ostomy self-care-related videos depicted what it is like living life with a stoma, and what self-care strategies were found to be effective and what were not. This type of data would be difficult to gain using traditional observational techniques and surveys.

End User Experience With Ostomy Medical Device Technology

The narrators who have undergone ostomies used a variety of supplies at different frequencies to change their ostomy

appliance and discussed challenges related to supplies, challenges with peristomal skin health, appliance adhesion, leakage, and not having control over the stoma output which reflects their user needs. For example, peristomal skin complications were identified by narrators who have undergone ostomies in approximately one-third of analyzed ostomy self-care YouTube videos, and stoma powder (used to treat skin complications) and skin treatments were included and discussed in 12 (34%) and 8 (23%) of the videos depicting an ostomy appliance change procedures, respectively. This finding aligns with extensive research on peristomal skin complications following an ostomy surgery representing a common challenge [30,37,38] and negatively impacting the quality of life of people who have undergone ostomies [39]. The use of peristomal skin protection supplies such as barrier rings, skin barriers, and ostomy paste in 17 (49%), 15 (43%), and 6 (17%) of videos, respectively, and adhesion supplies such as extenders, ostomy belts, and skin adhesives in 5 (14%), 4 (11%), and 3 (9%) of the videos, respectively, manifests the need for properly fitting and adhering external pouching system [17,40]. Thus, the selection of supplies may have been associated with attempts to address appliance adhesion and leakage issues and thus peristomal skin complications.

Narrators who have undergone ostomies reported usability issues with supplies, such as the difficulty with out-of-package baseplate not adhering to "warm" body temperature skin, supplies that are effective in cleaning but compromise peristomal skin health and ostomy appliance adhesion and cause pain (eg, alcohol-based ostomy paste), which call for design improvements of ostomy supplies or redesign related to their properties (eg, baseplate adhesion to peristomal skin, or alcohol-free supplies).

Moldable technology is believed to improve peristomal skin health, reduce the incidence of irritant dermatitis, and be well-perceived by people who have undergone ostomies for its ease of use, ease of learning, and comfort [41,42]. However, in our study, only 3 (9%) of videos depicting the full or part of the ostomy appliance change procedure included a moldable baseplate compared to 25 (71%) of videos where the baseplate was cut to fit or precut and people who have undergone ostomies reported challenges with cutting the baseplate to size. If the stoma hole is too small, the stoma can be damaged, if the stoma hole is too big, the output is more likely to leak on the peristomal skin which in turn will lead to peristomal skin complications. Generally, there are 3 designs of baseplate, precut (cut by supplier), cut-to-size (cut by a person who has undergone an ostomy), and moldable. Moldable baseplate technology does not require cutting the baseplate with scissors but rather using fingers to roll flexible material to size and around ("turtle necking") the stoma and allows for adjusting the stoma hole to fit snuggly around the stoma while also creating a protective barrier and preventing output from reaching peristomal skin [41]. The narrators who have undergone ostomies did not elaborate on the reason for cutting versus molding baseplate to size, except for 1 video where the narrator who has undergone an ostomy had her husband change the appliance and commented on the molding being "easier to use at first." Therefore, further studies should be conducted to explain this trend and to more specifically identify user needs. Additionally, future work should determine if people who have undergone ostomies are aware of and educated on the benefits of using moldable baseplate technology versus using cut-to-fit or precut appliances and if moldable appliances are widely available to people who have undergone ostomies.

Experts' User-Generated Ostomy Self-Care Strategies

The tips and tricks discussed by the narrators who have undergone ostomies in the videos we examined provide evidence of problem-solving strategies useful for decision-making and ostomy self-management. Ostomy self-care requires complex daily decision-making to solve ostomy-related challenges such as compromised appliance adhesion, leaks, or peristomal skin complications discussed above. People who have undergone ostomies have to monitor their daily output, diet, and stomal or peristomal skin health; perform ostomy appliance changes; and plan future self-management. In previous studies of professional expertise, experts stood out in their ability to identify problems [43], understand functional relationships [44], and use problem-solving strategies [45].

Lippa et al [46] suggest that in the self-management of chronic conditions, such as diabetes, professional expertise may be useful for aiding decision-making and supporting patient's strategies. Lippa et al [46] also found that diabetic patients who used functional knowledge and problem-solving strategies reported higher levels of adherence to treatment and glycemic control and that descriptive knowledge about diabetes may not necessarily correlate with effective self-management [46]. In our study, narrators who have undergone ostomies demonstrated their expert understanding while elaborating and showing their self-care strategies and adaptation to ostomy-related challenges. These strategies relate to supplies use and organization, ostomy appliance adhesion, baseplate sizing, maintenance of healthy peristomal skin, pouch pancaking and ballooning, and adjustments to lifestyle changes.

Finally, our study of ostomy self-care and end user experiences with ostomy appliance change technology complements the "Ostomy Life Study," [47] a global effort to gain knowledge about what is it like to live with a stoma and the user experience with ostomy-related supplies. Our study identifies an efficient alternative to traditional observational and survey methods for obtaining end user data that have implications for addressing patient safety and quality of care issues.

Limitations

Several limitations should be considered when interpreting the results of this study and guiding the direction of future research. In our sample of videos, there were more ileostomies than colostomies, and there were no urostomies, therefore, our findings did not compare different types of ostomies. Further, the narrators who have undergone ostomies were essentially "performing" and demonstrating for the video, therefore, these videos may not capture the routine procedures people who have undergone ostomies use for self-care. In our data analysis, half of the narrators who have undergone ostomies were within an assumed age of 20-30 years and approximately one-third of narrators who have undergone ostomies were in their middle

age. This is likely due to most YouTube users falling between the ages of 18 and 44 years [48].

There were also limitations related to demographics in this study. All videos analyzed were narrated in English, but the global representation is unknown. This may limit the global generalizability of this study. Additionally, we did not examine any race, gender, or ethnicity differences in the experiences of people who have undergone ostomies.

Future Work

Further, it should examine differences in ostomy appliance change procedure and self-care strategies across all the types of ostomies, more specifically, whether ileostomies are more prone to peristomal skin complications considering ileostomies' output is frequent, more liquid, and high in digestive enzymes [32]; and the prevalence and decision-making of using a moldable versus cut-to-fit appliance technology among people who have undergone ostomies. Future research should also specifically study younger and older people who have undergone ostomies and their experiences, complement this work in other languages and for individuals with experiences in other health care systems, and examine if there are differences among various populations of people who have undergone ostomies.

We identified 28 individual narrators who have undergone ostomies, and their videos were widely viewed. In addition to counting likes, future research should examine who is viewing these videos, what they are learning from them, and the content of viewer comments and discussions related to the video content. The comments analysis could provide a rich discussion around the strategies and ostomy self-care provided in the ostomy self-care videos.

In our study, we purposely did not include any video content from medical professionals or health care professional organizations because it can contain different information about ostomy self-care strategies than what is practiced by people who have undergone ostomies in home settings [35]. We intended to evaluate end user experiences with ostomy self-care rather than how people engage with YouTube videos created by health care professionals. However, we recognize there is value in comparing user-generated information to what clinicians and health-related government organizations may communicate about ostomy self-care as they use YouTube to disseminate health information [1] that is considered to be the most reliable and accurate information [6]. Future research should examine the content in the ostomy self-care videos created by health care professionals and professional organizations and evaluate how they compare to the user-generated ostomy self-care YouTube videos examined in this study.

Conclusions

This study used a novel approach for ostomy care (YouTube videos) to gain insights into ostomy self-care in home settings and end user interactions with medical devices while performing self-care, which are difficult to gather using traditional behavioral techniques. Using YouTube or other publicly available content created by individuals dealing with a chronic condition is a low-cost means to identify successful strategies, potential design problems, and potential workaround for issues that would be difficult to obtain in other study designs and approaches. The analysis of 65 videos showed that people who have undergone ostomies are willing to share their personal experience with ostomy self-care on the web and that these videos are viewed by the public. The findings also identify user needs related to medical device technology and indicate that narrators who have undergone ostomies were able to use problem-solving strategies for decision-making and successful ostomy self-management. This study demonstrates the potential value of these videos for filling the gap between health information delivered by health care professionals and the lived experience of individuals dealing with a chronic condition.

Additional information related to long-term ostomy self-care in a home (nonhospital) setting and the differences between what is clinically recommended and how the procedures and supplies are actually used by individuals living with ostomies needs to be more clearly understood for the design and testing of future ostomy supplies as well as for patient education from clinicians. Future research could use these approaches to address other types of clinical information that other clinical populations need to learn or be able to access when at home as well as strategies to design better medical devices for home health care in the future.

Acknowledgments

This work was supported by a Transformative Seed Grant from the Prisma Health System Research Division, a Center of Biomedical Research Excellence (SC TRIMH) funded by NIH/NIGMS (P20GM121342), and institutional funding from the Creative Inquiry Program at Clemson University.

Conflicts of Interest

None declared.

References

1. Gooding LF, Gregory D. Descriptive analysis of YouTube music therapy videos. J Music Ther 2011;48(3):357-369. [doi: 10.1093/jmt/48.3.357] [Medline: 22097103]
2. Steinberg PL, Wason S, Stern JM, Deters L, Kowal B, Seigne J. YouTube as source of prostate cancer information. Urology 2010 Mar;75(3):619-622. [doi: 10.1016/j.urology.2008.07.059] [Medline: 19815255]

3. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015 Sep;21(3):173-194 [FREE Full text] [doi: 10.1177/1460458213512220] [Medline: 24670899]
4. Ceci L. Statista. 2024 Apr. URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2025-02-19]
5. Altunisik E, Firat YE, Kiyak Keceli Y. Content and quality analysis of videos about multiple sclerosis on social media: the case of YouTube. Mult Scler Relat Disord 2022 Sep;65:104024. [doi: 10.1016/j.msard.2022.104024] [Medline: 35841755]
6. Basim P, Argun D. A Qualitative Analysis of Ostomy-Related Patient Education Videos on YouTube. Adv Skin Wound Care 2021 Jun 01;34(6):314-320. [doi: 10.1097/01.ASW.0000744340.11915.93] [Medline: 33979820]
7. Cetin A. Evaluation of YouTube video content related to the management of hypoglycemia. Cureus 2021 Jan 06;13(1):e12525 [FREE Full text] [doi: 10.7759/cureus.12525] [Medline: 33564529]
8. Ruppert L, Køster B, Siegert AM, Cop C, Boyers L, Karimkhani C, et al. YouTube as a source of health information: analysis of sun protection and skin cancer prevention related issues. Dermatol Online J 2017 Jan 15;23(1):1-10 [FREE Full text] [doi: 10.5070/D3231033669] [Medline: 28329465]
9. Keelan J, Pavri-Garcia V, Tomlinson G, Wilson K. YouTube as a source of information on immunization: a content analysis. JAMA 2007 Dec 05;298(21):2482-2484. [doi: 10.1001/jama.298.21.2482] [Medline: 18056901]
10. Frohlich DO, Zmyslinski-Seelig A. The presence of social support messages on YouTube videos about inflammatory bowel disease and ostomies. J Health Commun 2012;27(5):421-428. [doi: 10.1080/10410236.2011.606524] [Medline: 21962112]
11. Gómez-Zúñiga B, Fernandez-Luque L, Pousada M, Hernández-Encuentra E, Armayones M. ePatients on YouTube: analysis of four experiences from the patients' perspective. Med 2 0 2012;1(1):e1 [FREE Full text] [doi: 10.2196/med2.2039] [Medline: 25075229]
12. Tase A, Vadhwana B, Buckle P, Hanna GB. Usability challenges in the use of medical devices in the home environment: A systematic review of literature. Appl Ergon 2022 Sep;103:103769 [FREE Full text] [doi: 10.1016/j.apergo.2022.103769] [Medline: 35430424]
13. Dul J, Bruder R, Buckle P, Carayon P, Falzon P, Marras WS, et al. A strategy for human factors/ergonomics: developing the discipline and profession. Ergonomics 2012;55(4):377-395 [FREE Full text] [doi: 10.1080/00140139.2012.661087] [Medline: 22332611]
14. Bitkina OV, Kim HK, Park J. Usability and user experience of medical devices: An overview of the current state, analysis methodologies, and future challenges. Int J Ind Ergon 2020 Mar;76:102932 [FREE Full text] [doi: 10.1016/j.ergon.2020.102932]
15. Davis BR, Valente MA, Goldberg JE, Lightner AL, Feingold DL, Paquette IM, Prepared on behalf of the Clinical Practice Guidelines Committee of the American Society of ColonRectal Surgeons. The American Society of Colon and Rectal Surgeons clinical practice guidelines for ostomy surgery. Dis Colon Rectum 2022 Oct 01;65(10):1173-1190. [doi: 10.1097/DCR.0000000000002498] [Medline: 35616386]
16. United Ostomy Associations of America. United Ostomy Associations of America. New ostomy patient guide. 2024 Apr. URL: https://www.ostomy.org/wp-content/uploads/2024/04/UOAA-New-Ostomy-Patient-Guide-2024-04.pdf [accessed 2025-02-19]
17. Herlufsen P, Olsen AG, Carlsen B, Nybaek H, Karlsmark T, Laursen TN, et al. Study of peristomal skin disorders in patients with permanent stomas. Br J Nurs 2006;15(16):854-862. [doi: 10.12968/bjon.2006.15.16.21848] [Medline: 17108855]
18. Krouse RS, Grant M, McCorkle R, Wendel CS, Cobb MD, Tallman NJ, et al. A chronic care ostomy self-management program for cancer survivors. Psychooncology 2016 May;25(5):574-581 [FREE Full text] [doi: 10.1002/pon.4078] [Medline: 26804708]
19. Millard R, Cooper D, Boyle MJ. Improving self-care outcomes in ostomy patients via education and standardized discharge criteria. Home Healthc Now 2020;38(1):16-23. [doi: 10.1097/NHH.0000000000000816] [Medline: 31895893]
20. Prinz A, Colwell JC, Cross HH, Mantel J, Perkins J, Walker CA. Discharge planning for a patient with a new ostomy: best practice for clinicians. J Wound Ostomy Continence Nurs 2015;42(1):79-82. [doi: 10.1097/WON.0000000000000094] [Medline: 25333690]
21. Wound, Ostomy Continence Nurses Society Task Force. Wound, ostomy, and continence nursing: scope and standards of WOC practice, 2nd edition: an executive summary. J Wound Ostomy Continence Nurs 2018;45(4):369-387. [doi: 10.1097/WON.0000000000000438] [Medline: 29698290]
22. Gemmill R, Kravits K, Ortiz M, Anderson C, Lai L, Grant M. What do surgical oncology staff nurses know about colorectal cancer ostomy care? J Contin Educ Nurs 2011 Feb;42(2):81-88. [doi: 10.3928/00220124-20101101-04] [Medline: 21053790]
23. Tallman NJ, Cobb MD, Grant M, Wendel CS, Colwell J, Ercolano E, et al. Colostomy irrigation: issues most important to wound, ostomy and continence nurses. J Wound Ostomy Continence Nurs 2015;42(5):487-493. [doi: 10.1097/WON.0000000000000170] [Medline: 26336046]
24. Celik SS, Tuna Z, Yildirim M. The experience of urostomists who do not have access to pre‐operative and post‐operative stoma care nursing intervention. Int J of Uro Nursing 2014 Dec 29;9(2):101-107 [FREE Full text] [doi: 10.1111/ijun.12068]

25. Braumann C, Müller V, Knies M, Aufmesser B, Schwenk W, Koplin G. Quality of life and need for care in patients with an ostomy: a survey of 2647 patients of the Berlin Ostomy-Study (BOSS). Langenbecks Arch Surg 2016 Dec;401(8):1191-1201. [doi: 10.1007/s00423-016-1507-z] [Medline: 27659022]
26. Metcalf C. Stoma care: empowering patients through teaching practical skills. Br J Nurs 1999;8(9):593-600. [doi: 10.12968/bjon.1999.8.9.6621] [Medline: 10711005]
27. White T, Watts P, Morris M, Moss J. Virtual postoperative visits for new ostomates. Comput Inform Nurs 2019 Feb;37(2):73-79. [doi: 10.1097/CIN.0000000000000498] [Medline: 30562169]
28. Nichols TR, Inglese GW. The burden of peristomal skin complications on an ostomy population as assessed by health utility and the physical component summary of the SF-36v2. Value Health 2018 Jan;21(1):89-94 [FREE Full text] [doi: 10.1016/j.jval.2017.07.004] [Medline: 29304945]
29. Nybaek H, Knudsen DB, Laursen TN, Karlsmark T, Jemec GBE. Quality of life assessment among patients with peristomal skin disease. Eur J Gastroenterol Hepatol 2010 Feb;22(2):139-143. [doi: 10.1097/MEG.0b013e32832ca054] [Medline: 19918185]
30. Hendren S, Vu J, Suwanabol P, Kamdar N, Hardiman K. Hospital variation in readmissions and visits to the emergency department following ileostomy surgery. J Gastrointest Surg 2020 Nov;24(11):2602-2612 [FREE Full text] [doi: 10.1007/s11605-019-04407-6] [Medline: 31754986]
31. LeBlanc K, Whiteley I, McNichol L, Salvadalena G, Gray M. Peristomal medical adhesive-related skin injury: results of an international consensus meeting. J Wound Ostomy Continence Nurs 2019;46(2):125-136 [FREE Full text] [doi: 10.1097/WON.0000000000000513] [Medline: 30844869]
32. Lyon CC, Smith AJ, Griffiths CEM, Beck MH. The spectrum of skin disorders in abdominal stoma patients. Br J Dermatol 2000 Dec;143(6):1248-1260. [doi: 10.1046/j.1365-2133.2000.03896.x] [Medline: 11122029]
33. Salvadalena G. Incidence of complications of the stoma and peristomal skin among individuals with colostomy, ileostomy, and urostomy: a systematic review. J Wound Ostomy Continence Nurs 2008;35(6):596-607; quiz 608. [doi: 10.1097/01.WON.0000341473.86932.89] [Medline: 19018200]
34. Hill CE, Knox S. Essentials of consensual research. In: American Psychological Association. Washington, DC: American Psychological Association; 2021.
35. Silva BWACD, O LBD, Araújo AKD, Medeiros MBC, Melo VLD, Sena JFD, et al. Analysis of self-care videos on YouTube about exchange of intestinal ostomy bags. Rev Rene 2020;21:e44275. [doi: 10.15253/2175-6783.20202144275]
36. Yin R, Neyens DM. Online health resource use by individuals with inflammatory bowel disease: analysis using the national health interview survey. J Med Internet Res 2020 Sep 24;22(9):e15352 [FREE Full text] [doi: 10.2196/15352] [Medline: 32969831]
37. Sheetz KH, Waits SA, Krell RW, Morris AM, Englesbe MJ, Mullard A, et al. Complication rates of ostomy surgery are high and vary significantly between hospitals. Dis Colon Rectum 2014 May;57(5):632-637 [FREE Full text] [doi: 10.1097/DCR.0000000000000038] [Medline: 24819104]
38. Taneja C, Netsch D, Rolstad BS, Inglese G, Eaves D, Oster G. Risk and economic burden of peristomal skin complications following ostomy surgery. J Wound Ostomy Continence Nurs 2019;46(2):143-149 [FREE Full text] [doi: 10.1097/WON.0000000000000509] [Medline: 30844870]
39. Baxter NN, Novotny PJ, Jacobson T, Maidl LJ, Sloan J, Young-Fadok TM. A stoma quality of life scale. Dis Colon Rectum 2006 Feb;49(2):205-212. [doi: 10.1007/s10350-005-0275-6] [Medline: 16392025]
40. Arumugam PJ, Bevan L, Macdonald L, Watkins AJ, Morgan AR, Beynon J, et al. A prospective audit of stomas-analysis of risk factors and complications and their management. Colorectal Dis 2003 Jan;5(1):49-52. [doi: 10.1046/j.1463-1318.2003.00403.x] [Medline: 12780927]
41. Szewczyk MT, Majewska G, Cabral MV, Hölzel-Piontek K. The effects of using a moldable skin barrier on peristomal skin condition in persons with an ostomy: results of a prospective, observational, multinational study. Ostomy Wound Manage 2014 Dec;60(12):16-26 [FREE Full text] [Medline: 25485549]
42. Liu G, Chen Y, Luo J, Liu A, Tang X. The application of a moldable skin barrier in the self-care of elderly ostomy patients. Gastroenterol Nurs 2017;40(2):117-120. [doi: 10.1097/SGA.0000000000000143] [Medline: 26987102]
43. Dreyfus H, Drey-fus SE. Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer. New York: The Free Press; Jun 1987:110-111.
44. Chi MTH, Feltovich PJ, Glaser R. Categorization and representation of physics problems by experts and novices. Cognitive Science 1981;5(2):121-152. [doi: 10.1207/s15516709cog0502_2]
45. Klein G. Sources of Power: How people make decisions. United States: The MIT Press; 1999.
46. Lippa KD, Klein HA, Shalin VL. Everyday expertise: cognitive demands in diabetes self-management. Hum Factors 2008 Feb;50(1):112-120. [doi: 10.1518/001872008X250601] [Medline: 18354975]
47. Coloplast. Ostomy Life Study. Coloplast Corp. URL: https://www.coloplast.us/landing-pages/ols [accessed 2025-02-19]
48. GMI Research Team. GLOBAL MEDIA INSIGHT - Dubai Digital Interactive Agency. URL: https://www.globalmediainsight.com/blog/youtubeusersstatistics/#YouTube_Users_by_Year [accessed 2025-02-19]

Edited by A Kushniruk; submitted 18.09.24; peer-reviewed by C Milne, C Ruiz-Nunez; comments to author 07.11.24; revised version received 06.12.24; accepted 20.12.24; published 06.03.25.

Please cite as:
Haughey M, Neyens DM, Hopkins CS, Gonzaga C, Harman M
Identifying Strategies for Home Management of Ostomy Care: Content Analysis of YouTube
JMIR Hum Factors 2025;12:e66634
URL: https://humanfactors.jmir.org/2025/1/e66634
doi:10.2196/66634
PMID:40053741

Original Paper

Development of a Patient-Centered Symptom-Reporting Application in Pharmacy Settings Using a Hierarchical Patient-Friendly Symptom List: Developmental and Usability Study

Seiya Watanabe[1], BSc; Hayato Kizaki[1], PhD; Satoko Hori[1], PhD

Division of Drug Informatics, Keio University Faculty of Pharmacy, Tokyo, Japan

Corresponding Author:
Satoko Hori, PhD
Division of Drug Informatics
Keio University Faculty of Pharmacy
1-5-30 Shibakoen, Minato-ku
Tokyo, 105-8512
Japan
Phone: 81 354002650
Email: satokoh@keio.jp

Abstract

Background: Effective symptom identification, a key responsibility for community pharmacists, requires patients to describe their symptoms accurately and comprehensively. However, current practices in pharmacies may be insufficient in capturing patient-reported symptoms comprehensively, potentially affecting the quality of pharmaceutical care and patient safety.

Objective: This study aimed to construct a new, hierarchical symptom list derived from the Patient-Friendly Term List of the Medical Dictionary for Regulatory Activities (MedDRA) and to develop and evaluate a mobile app incorporating this list for facilitating symptom reporting by patients in pharmacy settings. The study also aimed to assess the usability and acceptance of this app among potential users.

Methods: Subjective symptom-related terms were extracted from the Patient-Friendly Term List version 23.0 of the MedDRA. These terms were systematically consolidated and organized into a hierarchical, user-friendly symptom list. A mobile app incorporating this list was developed for pharmacy settings, featuring a symptom selection interface and a free-text input field for additional symptoms. The app included an instructional video explaining the importance of symptom reporting and guidance on navigation. Usability tests and semistructured interviews were conducted with participants aged >20 years. Interview transcripts were analyzed using the Unified Theory of Acceptance and Use of Technology (UTAUT) model to evaluate factors influencing the acceptance of technology.

Results: From the initial 1440 terms in the Patient-Friendly Term List, 795 relevant terms were selected and organized into 40 site-specific subcategories, which were then grouped into broader site categories (mental, head, trunk, upper limb, lower limb, physical condition, and others). These terms were further consolidated into 211 patient-friendly symptom terms, forming a hierarchical symptom list. The app's interface design limited options to 10 items per screen to assist with decision-making. A total of 5 adults participated in the usability test. Participants found the interface intuitive and easy to use, requiring minimal effort, and provided positive feedback regarding the potential utility of the app in pharmacy settings. The UTAUT analysis identified several facilitating factors, including ease of use and the potential for enhanced pharmacist-patient communication. However, concerns were raised about usability for older adults and the need for simplified technical terminology.

Conclusions: The user-friendly app with a hierarchically structured symptom list and complementary free-text entry has potential benefits for improving the accuracy and efficiency of symptom reporting in pharmacy settings. The positive user acceptance and identified areas for improvement provide a foundation for further development and implementation of this technology to enhance communication between patients and pharmacists. Future improvements should focus on addressing usability for older adults and simplifying technical terminology.

(JMIR Hum Factors 2025;12:e71439) doi:10.2196/71439

KEYWORDS

patient symptom monitoring; hierarchical symptom list; community pharmacy; interview survey; mobile application

Introduction

Patient-reported symptoms are critical indicators in health care, essential for maintaining patient safety and improving therapeutic outcomes in clinical settings [1,2]. In Japan, the Ministry of Health, Labour and Welfare launched "Pharmacy Visions for Patients" in 2015, clearly outlining the role of community pharmacies, including comprehensive monitoring of medications, as well as providing tailored drug management and guidance [3]. A key responsibility for community pharmacies in this context is to gather and monitor patient-reported symptoms, essential for identifying potential health issues and ensuring timely intervention.

For pharmacists, collecting patient-reported symptoms plays multiple critical roles in medication therapy management. First, it facilitates the early detection and monitoring of potential adverse drug reactions, enabling timely interventions to prevent serious complications. Second, comprehensive symptom information helps pharmacists assess therapeutic effectiveness and adjust medication regimens accordingly. Third, systematic symptom monitoring supports pharmacists in providing targeted patient education and improving medication adherence through a better understanding of patient experiences. These activities are fundamental to the pharmacist's role in ensuring medication safety and optimizing therapeutic outcomes.

While health care providers traditionally collect symptom information through direct questioning, a patient-centered approach that enables patients to report symptoms at their own pace using familiar terminology may improve the accuracy and comprehensiveness of symptom reporting. Such an approach acknowledges patients as active participants in their health care, potentially leading to better identification of health-related issues and more effective interventions. A cross-sectional study demonstrated that structured symptom reporting tools can help identify numerous patient-reported symptoms and their potential associations with medications, providing valuable information for medication reviews [4]. However, another previous study showed that patients often underreported symptoms to health care providers, either because they do not attribute symptoms to the medication or do not recognize the significance of the symptoms [5].

To facilitate patient reporting, various symptom questionnaires have been developed, enabling patients to describe their experiences [6]. However, many of these tools lack thorough validation. In Japan, Nojo et al [7] introduced the "Adverse Drug Reaction Signal Check Sheet," which lists subjective symptoms of side effects for patients taking high-risk drugs. Although limited to specific drugs, the sheet proved effective in prompting patients to communicate their symptoms during consultations with pharmacists. This approach suggests that selecting symptoms from a structured list can support patient reporting.

The Medical Dictionary for Regulatory Activities (MedDRA), a globally recognized dictionary of medical terminology, includes a Patient-Friendly Term List that reflects a wide range of symptoms reported by patients in safety databases [8]. MedDRA has a hierarchical structure with organ categories at the top, and its Patient-Friendly Term List is a supplementary list made up of the lowest-level terms of the MedDRA. This structured approach allows patients to progressively identify and report their symptoms. However, international research has indicated that patients do not fully use this list when reporting adverse drug reactions [9], suggesting challenges in the direct adoption of a symptom-reporting system.

This study aimed to construct a new hierarchical list of symptoms derived from the Patient-Friendly Term List of the MedDRA to simplify and facilitate incremental symptom selection by patients. Moreover, we aimed to develop an app incorporating this list to enhance communications between patients and pharmacists by facilitating more accurate symptom reporting.

Methods

Creation of the New Hierarchical, Patient-Friendly List of Symptoms

A hierarchical list of subjective symptoms was created using the Patient-Friendly Term List version 23.0 from the MedDRA. MedDRA, developed by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, is a standardized medical terminology used in pharmaceutical regulation. The MedDRA Patient-Friendly Term List was selected for this study due to its standardized structure and global adoption in pharmacovigilance. This list consists of patient-friendly terms derived from MedDRA, ensuring that the terminology is both medically accurate and comprehensible to nonexpert users. The structured format of this list was expected to facilitate consistent symptom categorization while remaining accessible to patients, supporting pharmacists in identifying potential adverse drug reactions (ADRs) and optimizing medication therapy management.

Terms associated with subjective symptoms were extracted from the Patient-Friendly Term List and categorized by body part. Similar terms were consolidated into "symptoms," forming the lowest level of a 3-tier hierarchical structure, with "broader site categories" and "site-specific subcategories" as the upper levels. This new hierarchical list was reviewed and refined through multiple rounds of consensus by three researchers (WS, KH, and HS). WS was an undergraduate student in the Faculty of Pharmaceutical Sciences, and KH and HS were researchers in the Faculty of Pharmaceutical Sciences.

Development of a Patient Symptom-Reporting Application

An app was developed to incorporate the newly created hierarchical list of symptoms. The app featured a symptom selection interface, an instructional video explaining the importance of symptom reporting, and guidance on how to navigate the app. To assist with decision-making, no more than 10 options were displayed on the screen at a time. In addition to the structured symptom list, the app included a complementary free-text entry, allowing users to input symptoms that may not be represented on the list or when the listed symptoms did not fit their condition. The app is intended for use in a pharmacy setting and was designed to capture a

wide range of symptoms, beyond those related to side effects of specific drugs. Apache Cordova, an open-source mobile app development framework, was used for the development. The operating system used was Android, and the app was run on a HUAWEI MediaPad T3. The content of the app was written in Japanese.

Usability Study and Interviews

A usability study was conducted from September to October 2020 with adults aged >20 years, recruited through research flyers. Participants first completed a questionnaire on their demographics, medication use, history of ADRs, and familiarity with digital devices. They were then briefed on the purpose of the app in a pharmacy setting and the interview format. Participants tested the app on a HUAWEI MediaPad T3, entering personal and symptom data on 5 hypothetical patient scenarios (Table S1 in Multimedia Appendix 1). Following the app trial, a semistructured interview was conducted, based on the interview guide (Table S2 in Multimedia Appendix 1), to gather feedback on the app. Interviews were conducted in Japanese, focusing on impressions of the app from participants and their awareness of reporting side effects. All interviews were recorded with the consent of participants and conducted by a single researcher (WS).

Analysis of the Interview

Thematic analysis was applied to the recorded interview transcripts using a deductive and theoretical approach. The Unified Theory of Acceptance and Use of Technology (UTAUT) model was used for this analysis. Initially proposed by Venkatesh et al [10] in 2003, the UTAUT model has been widely adopted in the medical field in recent years [11]. This model is particularly useful for understanding factors influencing the adoption of new technologies. It posits that performance expectancy (expected benefits of using new technology), effort expectancy (expectations of the ease of use and understanding of new technology), and social influence (how much the user's decisions about technology are influenced by others) determine behavioral intention to adopt the technology. Behavioral intention, along with facilitating conditions, influence actual technology use. In addition, gender, age, experience with similar technologies, and spontaneity of use serve as general adjustment variables (moderators). The translation of these constructs was based on the work of Ono [12]. In total, 3 researchers conducted analyses to ensure objectivity.

Ethical Considerations

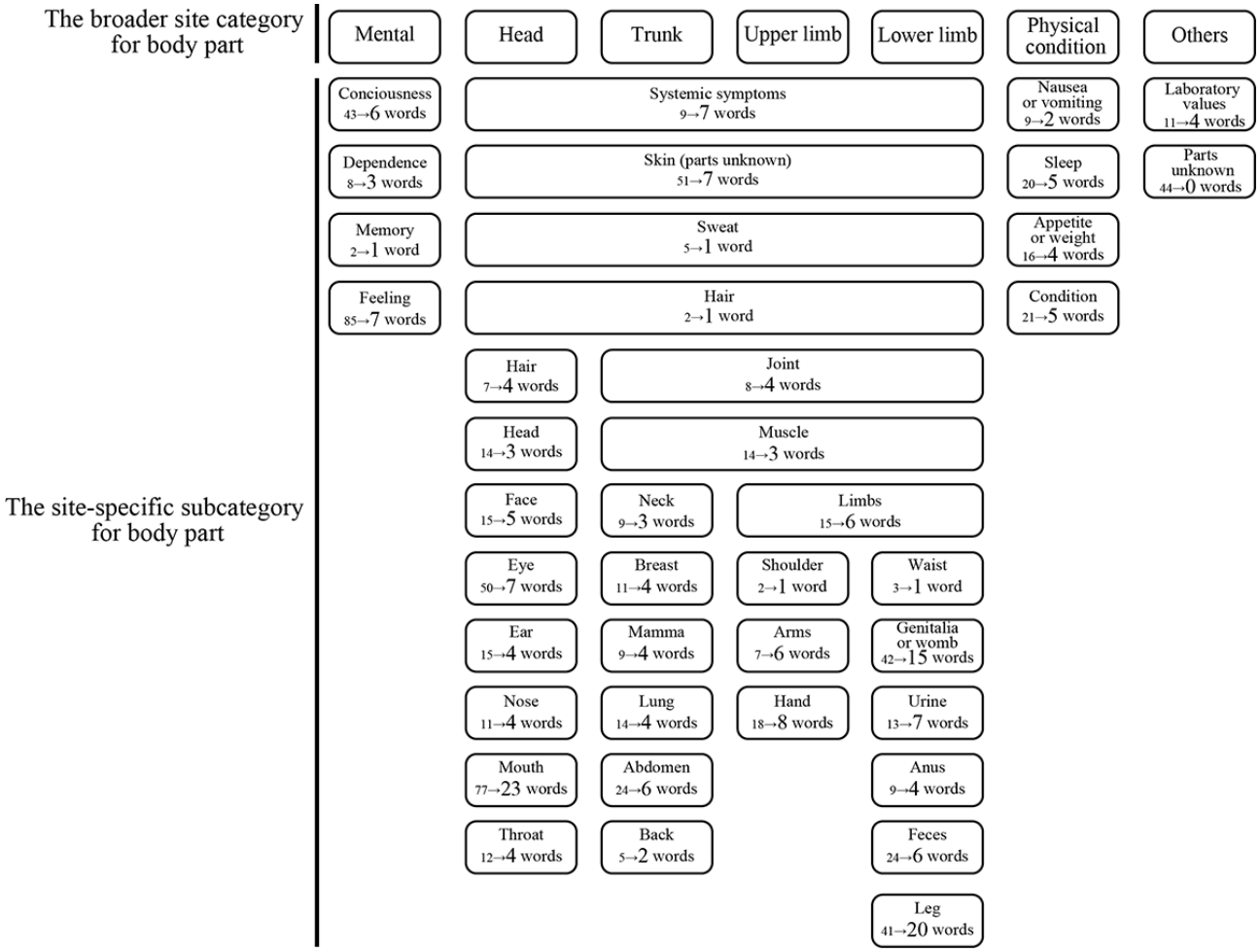
On the day of the usability study and interview, the survey content, privacy protection, and plans for publishing the research results were comprehensively explained to the participants, and written informed consent to participate in the survey was obtained. The preinterview questionnaire was self-administered anonymously. Participants received compensation in the form of a gift card (Quo Card) valued at 1000 JPY for their participation in the study. This survey was conducted after being approved by the Ethics Committee for Research Involving Human Subjects, Keio University Faculty of pharmacy, following the Ethical Guidelines for Medical Research Involving Human Subjects (200213-3).

Results

Creation of the New Patient-Friendly Symptom List

From the 1440 terms in the Patient-Friendly Term List version 23.0, duplicates with the same Japanese translations were removed, resulting in 1288 terms. Subsequently, 12 product-related terms such as "suspected counterfeit products" were excluded. In addition, 450 terms that were difficult for patients to self-identify, including those requiring a medical diagnosis or those specific to certain conditions, were removed. Another 31 terms unrelated to medications were deleted, leaving a total of 795 terms. The 795 terms were systematically sorted into 40 site-specific subcategories based on information related to the locus of impact of each word. These site-specific subcategories were grouped into 7 broad site categories: mental, head, trunk, upper limb, lower limb, physical condition, and others (Figure 1). Within each site-specific subcategory, similar terms were consolidated by grouping related symptoms (eg, various types of arm discomfort such as "arm paralysis" and "arm discomfort") and creating standardized descriptive names (eg, "tingling and discomfort in the arm") that accurately represent the consolidated symptoms. This consolidation process aimed to simplify symptom selection while maintaining clinical relevance, yielding 211 subjective symptoms (Figure 1). Finally, a hierarchical list of patient-friendly symptoms was created using the hierarchical structure of site categories, subcategories, and symptoms.

Figure 1. The broader site categories and site-specific categories for body parts, as well as the word counts before and after consolidation. The 795 terms were systematically sorted into 40 site-specific subcategories based on information related to the locus of impact of each word. These site-specific subcategories were grouped into 7 broad site categories: mental, head, trunk, upper limb, lower limb, physical condition, and others. Within each site-specific subcategory, similar terms were consolidated by grouping related symptoms (eg, various types of arm discomfort such as "arm paralysis" and "arm discomfort") and creating standardized descriptive names (eg, "tingling and discomfort in the arm") that accurately represent the consolidated symptoms. This consolidation process aimed to simplify symptom selection while maintaining clinical relevance. In this figure, the number on the left side of each arrow represents the count of terms before consolidation and the number on the right indicates the count following this unification process.



To integrate this list into the app, we designed a question flow that allowed for easy symptom selection. Broader and site-specific subcategories were modified and recombined into major site categories and site subcategories for an intuitive question flow (Figure S1 in Multimedia Appendix 1). As an example, the hierarchy under the major site category "upper limbs" and its subcategories is detailed in Table 1. The app included a free-text entry field labeled as "other" at each categorical level to capture symptoms that users could not find in the structured list. This feature was implemented to ensure comprehensive symptom reporting and to collect data for future improvements of the symptom list.

Table 1. Details of selection items (excerpt for upper limbs only). The major site categories, site subcategories, lower-level terms, and symptoms after consolidation were described in Japanese. The table presented here has been translated into English for this publication.

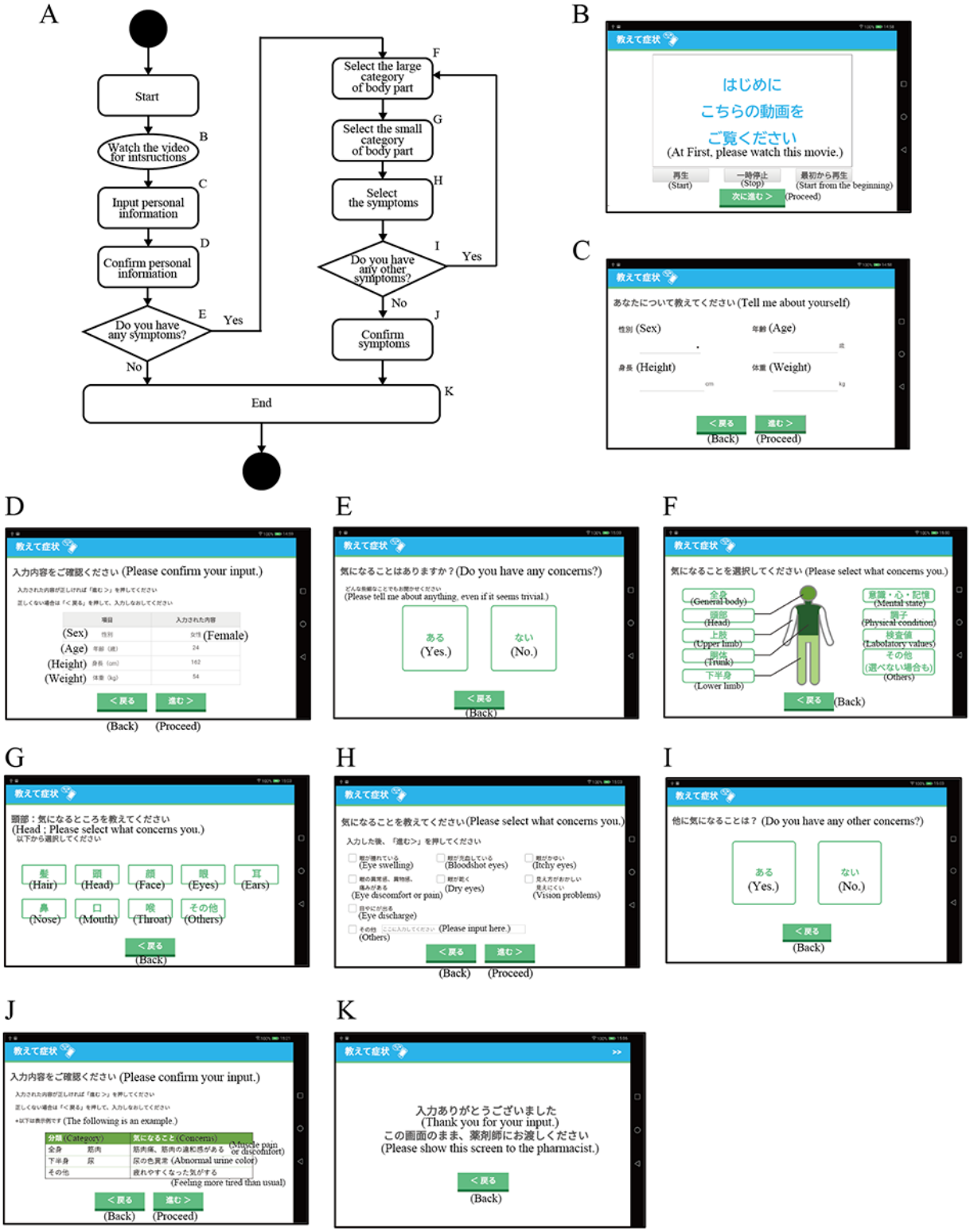
Major site categories and subcategories of body parts (LLT[a])	LLT code	Symptoms after consolidation
Upper limb		
Shoulder		
Scapular pain	10040610	Shoulder pain
Shoulder pain	10040617	Shoulder pain
Arm		
Upper limb pain	10033421	Arm pain
Elbow pain	10033424	Elbow pain
Arm rash	10037875	Arm rash
Arm paralysis	10003098	Tingling and discomfort in the arm
Arm discomfort	10049877	Tingling and discomfort in the arm
Arm swelling	10042680	Arm swelling and edema
Arm weakness	10050379	Weakness in the arm
Hand		
Brittle nails	10006373	Abnormal nails
Nail discoloration	10028691	Abnormal nails
Finger swelling	10042694	Swelling and edema in fingers
Hand swelling	10042695	Swelling and edema in fingers
Hand rash	10019117	Hand rash
Cold hands	10009860	Cold sensation in hands
Finger pain	10033428	Hand pain
Hand pain	10033430	Hand pain
Itching of both hands	10023087	Itchy hands
Wrist pain	10048692	Wrist pain
Clumsiness	10009696	Tingling, trembling, and discomfort in fingers
Hand cramp	10011287	Tingling, trembling, and discomfort in fingers
Finger deformity	10061156	Tingling, trembling, and discomfort in fingers
Stiffness of fingers	10016695	Tingling, trembling, and discomfort in fingers
Tingling in fingers	10029837	Tingling, trembling, and discomfort in fingers
Tingling in hands	10049681	Tingling, trembling, and discomfort in fingers
Reduced dexterity	10067727	Tingling, trembling, and discomfort in fingers
Hand tremor	10040530	Tingling, trembling, and discomfort in fingers

[a]LLT: lowest-level term.

Development of a Symptom-Reporting App

The screen transition diagram and display screens for the app are shown in Figure 2. Figure 2A shows the screen transition diagram of the app. The starting screen featured an instructional video explaining the purpose of the app and the importance of comprehensive symptom reporting (Figure 2B), followed by a series of intuitive selections for symptom identification (Figure 2E-J). The closed screen signaled the end of the session, prompting the user to pass the device on to the pharmacist (Figure 2K). To minimize cognitive overload, no more than 10 options were displayed per screen, making symptom selection easier and more manageable. An instructional video was included to explain the importance of symptom reporting, providing users with essential background knowledge before using the app. In addition, a free-text entry field was implemented to allow patients to describe symptoms not covered by the structured list, ensuring flexibility while maintaining structured reporting as the primary method. These features were designed to enhance the accuracy and completeness of symptom reporting, thereby improving communication between patients and pharmacists.

Figure 2. Screen transition diagram and display screen of the app. (A) Screen transition. (B) Start and video viewing section. (C) Input of personal information, (D) Confirmation of personal information. (E) Selection of the presence or absence of symptoms. (F) The large category of body parts. (G) The small category of body parts. (H) The selection of symptoms. (I) Selection of presence of other symptoms. (J) Confirmation of the selected symptoms. (K) End. The app screen in Japanese is shown since the actual app is in Japanese. English translations of the screen elements are provided in parentheses within the figure.



Usability of the Symptom-Reporting App

In total, 5 adults participated in the usability study and interviews. Table 2 presents the demographic characteristics of the interviewees. Overall, 4 of the 5 interviewees regularly used smartphones, and 4 of the 5 interviewees had previously experienced side effects. The interviews elicited valuable opinions on the usability of the app, as summarized in Table 3, with an analysis aligned with the UTAUT model presented in Figure S2 in Multimedia Appendix 1.

Table 2. Basic information of interviewees.

No.	Age (years)	Sex	Occupation	Types of medicines taken, n	Experience of side effects	Smartphone or tablet	
						Possession	Frequency
1	79	Female	Unemployed	9	Yes	No	Do not use
2	24	Male	Student	3	Yes	Yes	Frequently
3	23	Male	Office worker	None	No	Yes	Frequently
4	50	Male	Sole proprietorship	3	Yes	Yes	Frequently
5	54	Female	Medical office worker	1	Yes	Yes	Frequently

Table 3. Opinions on app usability based on the Unified Theory of Acceptance and Use of Technology (UTAUT) model.

UTAUT category	Details
Performance expectancy	
Ease of consultation	• Easier to answer than to be asked • Easier to tell something than in person • Gives a chance to talk • Unclear what the app is used for
Contribution for medication	• May contribute to the treatment of others
Feedback from health care professionals	• Good to have an opinion from a pharmacist • Good to have continuous monitoring • Good to have topics related to daily life • Good to have a comfortable relationship with the pharmacist
Other functions	• Good to provide information about medication • Good to enter one's medication status • Good to be able to discuss other issues besides symptoms • Good to be linked to other systems • Good to be used outside the pharmacy • Concerns about adding too many functions
Effort expectancy	
Ease of use	• Usable without problems • Easy to use selective forms • Easy to be guided by selective forms • Negative feelings though none of the options apply to me • Difficult to enter free-text fields • Difficult to check, change, or add information • Difficult to use for older people who are not used to using smartphones or tablets
Expression of question items	• Usable without problems • Very detailed items • Many specialized words and phrases • Difficult to understand the words of the site classification • Difficult to understand the way the questions are asked
Social influence	
Influence from health care professionals	• Easy to answer if a health care provider asks about symptoms • App that can be used by pharmacists in medication instruction
Information collection and management	• Resistance to data being stored
Facilitating conditions	
Appearance	• Easy-to-read layout • Easy-to-read colors • Easy-to-read font size • Difficult reading text for older people
Device	• Difficult to use keyboard • No problems with the size of device • Easy to use the device with horizontal screen display • Probably being too small for older people
Behavioral intention	
Use of app	• Want to use the app • Should be used also outside of pharmacies • Probably not being able to be used for older people alone • Sometimes difficult to tell something even with app

In terms of "performance expectancy," participants expressed a preference for using the app to consult with pharmacists because "it is easier to answer than when being asked by a person," "it is easier to communicate than doing it face-to-face,"

and “it gives me a chance to talk to someone.” Some participants suggested that “it would be beneficial to receive expert opinions as feedback from a pharmacist.” In terms of “effort expectancy” (expectations about the ease of use and understanding of new technology), usability and expressions were both deemed “usable without problems,” while issues such as “too many technical terms” and “I feel bad when there are no applicable choices” were identified. Regarding “social influence” (the extent to which the user’s technology decisions are influenced by others), opinions included “it is easy to answer when a medical professional asks about symptoms” and “pharmacists can use it in medication guidance,” indicating that the influence of medical professionals contributes to the behavioral intention to use the technology. Regarding the “facilitating conditions,” it was suggested that there should be no problems with the design and terminal. Based on the above, all participants were positive about their behavioral intention to use the app, and some suggested that the app should be used outside pharmacies. Concerns were raised regarding the use of the app by older adults.

Several factors that did not align with the UTAUT model but significantly affected how ADRs reported to health care providers were identified. These factors included “I recognize pharmacist as a person to consult, but do not want to consult,” “I will not consult if not sure of side effect,” “I will not consult if the side effect is minor or not necessary to discuss by myself,” and “I will not consult if I think it is a side effect.” In addition, factors specifically affecting the reporting of ADRs to health care providers comprised “I recognize that it is an ADR but do not want to discuss it with a pharmacist,” “I do not discuss it unless I am sure it is an adverse drug reaction,” “I do not discuss it unless I am sure it is an adverse drug reaction,” and “I feel uncomfortable reporting an adverse drug reaction.”

Discussion

Principal Findings

In this study, we developed a hierarchical list of symptoms based on the MedDRA Patient-Friendly Term List and created an app that enables patients to easily select symptoms. The potential use of the app in pharmacies was well received, indicating that it could significantly aid patients in communicating their symptoms more effectively.

Our newly created list comprised 211 symptoms, and to improve user-friendliness, we implemented a “questionnaire flow” within the app, based on the hierarchical structure. This approach allows users to systematically select symptoms by navigating through major site categories, site subcategories, and detailed symptoms. Participants using the app in 5 hypothetical patient scenarios praised its ease of use, confirming the feasibility of using the hierarchical list of symptoms and the app for effective symptom selection. Previous research revealed that approximately 20% of reports on side effects used patient-friendly terms, with the majority opting for free-text entries [5]. By limiting displayed options to 10 or fewer, while providing a free-text input field for unlisted symptoms, our app may improve the ease of symptom identification. This design feature was further supported by the overall positive feedback from the participants. Future updates to the symptom list, informed by the analysis of real-world, free-text entries, could further streamline symptom reporting.

Participants in our study expressed their opinions about reporting usual self-aware symptoms such as “I do not discuss subjective symptoms unless I am sure they are side effects” and “I do not mention side effects if they seem minor or unnecessary to report,” which aligns with reasons reported in previous research for underreporting symptoms to health care providers [5]. These responses highlight common barriers to symptom reporting in health care settings, including patients’ uncertainty about whether symptoms are drug-related and their hesitation to mention symptoms they perceive as minor. To address these universal challenges, our app was designed with a structured interface and a hierarchical organization of symptoms, guiding patients through a systematic symptom-reporting process. The hierarchical structure of the app enables patients to progressively identify and select symptoms, potentially increasing the comprehensiveness of reported symptoms while reducing psychological barriers to reporting even minor or uncertain symptoms. Positive feedback from participants, such as “easier to answer than being asked by someone,” “easier to communicate than doing it face-to-face,” and “becomes a trigger to talk,” suggests that this structured approach could effectively address these reporting barriers. In addition, participants’ feedback indicating the app’s effortlessness (“can use without a problem”) and willingness to use (“I want to use it”) suggests its potential for successful implementation in pharmacy practice.

From the perspective of health care providers, pharmacists can review these structured patient reports before medicine consultation, enabling more focused and efficient discussions. These features, combined with the ability of the app to document patient-reported outcomes systematically, provide a foundation for continuous monitoring of therapy effectiveness and early detection of drug-related problems.

While our usability study received positive feedback regarding the potential of the app, we also identified several usability issues, particularly among older users. A participant aged 79 years expressed concerns, stating “I might not be able to use it on my own,” and other feedback pointed to specific issues such as “The size may be too small for elderly users” and “The text might be difficult to read.” Given that many customers of pharmacies are older adults, future usability assessments should focus on this population.

Implementing our app in pharmacy settings strategically aligns with the evolving health care landscape in Japan, particularly in aging societies where pharmacies are expanding beyond traditional medication dispensing [3,13]. Pharmacies offer several unique advantages as ideal implementation sites, including their high accessibility, regular patient visits, and prescription preparation waiting time, which offers a natural opportunity for symptom reporting without disrupting workflow. While the app could be valuable in various health care settings, including clinics and hospitals, pharmacies offer unique advantages. Pharmacists’ expertise in medication, combined with their frequent patient interactions, makes them ideal providers for continuous symptom monitoring and self-care

support. Their role has evolved from traditional medication dispensing and adverse effect monitoring to more comprehensive patient care support. Integrating the app into pharmacy practice could particularly support patient engagement in their own health care by enabling systematic symptom reporting in a familiar, low-pressure environment. This approach not only supports pharmacists' evolving role in comprehensive patient care but also promotes active patient participation in medication management.

Future research should systematically evaluate the effectiveness of the app in achieving its intended outcomes. Key areas of investigation include the impact of the app on the number and types of symptoms reported. Quantitative measurements will be valuable in assessing the impact on pharmaceutical interventions, including the identification of drug-related problems and subsequent care recommendations. Further research opportunities lie in evaluating improvements in the quality of patient-pharmacist communication and the effect of the app on the efficiency of pharmacy workflow. These investigations would benefit from including a larger, more diverse patient population in real-world pharmacy settings, with particular attention to older adults who represent a significant portion of pharmacy patients.

Although our sample size of 5 participants is consistent with recommendations for initial usability testing suggesting that 5 users can identify approximately 80% of major usability issues [14], this small sample size limits the generalizability of our findings. This limitation highlights the need for additional testing with a larger and more diverse user group to ensure the app meets the needs of all potential users in pharmacy settings.

Conclusions

We successfully developed a new hierarchical list of symptoms and an accompanying app. The app, designed to enable users to select symptoms from a structured list, was positively accepted and showed strong potential for improving patient behavior in symptom reporting. Future improvements to the design of the app design, particularly for older users, will further enhance its utility in both pharmacy settings and broader health care environments.

Acknowledgments

We would like to thank all the participants in this study.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to privacy or ethical restrictions but are available from the corresponding author on reasonable request.

Authors' Contributions

WS contributed to conceptualization, methodology, software, investigation, formal analysis, validation, and writing—review and editing. KH handled conceptualization, methodology, project administration, validation, visualization, writing—original draft, and writing—review and editing. HS performed conceptualization, methodology, project administration, supervision, validation, writing—original draft, and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials including hierarchical structure of symptom selection categories, UTAUT model application results, model patient profiles, and interview guide. UTAUT: Unified Theory of Acceptance and Use of Technology.
[DOCX File , 162 KB - humanfactors_v12i1e71439_app1.docx]

References

1. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016;34(6):557-565. [doi: 10.1200/jco.2015.63.0830]
2. Lizán L, Pérez-Carbonell L, Comellas M. Additional value of patient-reported symptom monitoring in cancer care: a systematic review of the literature. Cancers (Basel) 2021;13(18):4615 [FREE Full text] [doi: 10.3390/cancers13184615] [Medline: 34572842]
3. "Pharmacy Vision for Patients" - From "Front of the Door" to "Family Care" and then to "Community" - has been formulated. Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/houdou/0000102179.html [accessed 2024-08-28]
4. Schoenmakers TWA, Teichert M, Wensing M, de Smet PAGM. Evaluation of potentially drug-related patient-reported common symptoms assessed during clinical medication reviews: a cross-sectional observational study. Drug Saf 2017;40(5):419-430 [FREE Full text] [doi: 10.1007/s40264-017-0504-7] [Medline: 28205099]

5. Martin ML, Chung H, Rydén A. Willingness to report treatment-related symptoms of immunotherapy among patients with non-small cell lung cancer. Qual Life Res 2022;31(4):1147-1155 [FREE Full text] [doi: 10.1007/s11136-021-02966-3] [Medline: 34383226]
6. Lim R, Ellett LK, Roughead EE, Cheah PY, Masnoon N. Patient-reported questionnaires to identify adverse drug reactions: a systematic review. Int J Environ Res Public Health 2021;18(22):11877 [FREE Full text] [doi: 10.3390/ijerph182211877] [Medline: 34831635]
7. Nojo Y, Yamanaka S, Son N, Koda K, Kitagaki K, Narukawa M, et al. Improving data collection in pharmacies on subjective symptoms experienced by patients on high-risk medication using "Adverse Drug Reaction Signal Check Sheet" [Article in Japanese]. Journal of Regulatory Science (RSMP) 2022 Sep;12(3):259-270. [doi: 10.14982/rsmp.12.259]
8. Patient-friendly term list. MedDRA. URL: https://www.meddra.org/patient-friendly-term-list [accessed 2025-01-09]
9. de Vries ST, Harrison J, Revelle P, Ptaszynska-Neophytou A, Radecka A, Ragunathan G, et al. Use of a patient-friendly terms list in the adverse drug reaction report form: a database study. Drug Saf 2019;42(7):881-886 [FREE Full text] [doi: 10.1007/s40264-019-00800-x] [Medline: 30725337]
10. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
11. Kim S, Lee KH, Hwang H, Yoo S. Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the Unified Theory of Acceptance and Use of Technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 2016;16:12 [FREE Full text] [doi: 10.1186/s12911-016-0249-8] [Medline: 26831123]
12. Ono A. A consumer theory of new technology acceptance [Article in Japanese]. Mita Business Review 2008 Apr;51(1):1-18 [FREE Full text]
13. Kawahara A, Nakai K, Kurata N, Kamei M. New role of Japanese pharmacists focusing on medication therapy management in a team care approach. J Pharm Pract Res 2020;50(1):112-113. [doi: 10.1002/jppr.1616]
14. Nielsen J, Landauer TK. A mathematical model of the finding of usability problems. 1993 Presented at: CHI '93: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems; April 24-29, 1993; Amsterdam, the Netherlands. [doi: 10.1145/169059.169166]

Abbreviations

ADR: adverse drug reaction
MedDRA: Medical Dictionary for Regulatory Activities
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Kushniruk, G Eysenbach; submitted 17.01.25; peer-reviewed by H Hosono, H Okada; comments to author 30.01.25; revised version received 07.02.25; accepted 10.02.25; published 06.03.25.

Please cite as:
Watanabe S, Kizaki H, Hori S
Development of a Patient-Centered Symptom-Reporting Application in Pharmacy Settings Using a Hierarchical Patient-Friendly Symptom List: Developmental and Usability Study
JMIR Hum Factors 2025;12:e71439
URL: https://humanfactors.jmir.org/2025/1/e71439
doi:10.2196/71439
PMID:40053749

Original Paper

Illustrating User Needs for eHealth With Experience Map: Interview Study With Chronic Kidney Disease Patients

Paula Valkonen[1], MA (Art and Des); Sini Hölsä[1], MSc (Tech); Johanna Viitanen[1], MSc (Tech), LicSc (Tech), DSc (Tech); Sini Leinonen[1], MSc (Tech); Nina Karisalmi[1], MSc (Tech); Virpi Rauta[2,3], MD, PhD

[1]Department of Computer Science, Aalto University, Espoo, Finland

[2]Department of Nephrology, University of Helsinki, Helsinki University Central Hospital, Helsinki, Finland

[3]Helsinki University Hospital, Abdominal Center, Nephrology, Helsinki, Finland

Corresponding Author:
Paula Valkonen, MA (Art and Des)
Department of Computer Science
Aalto University
P.O. Box 15400
Espoo, 00076
Finland
Phone: 358 947001
Email: paula.valkonen@aalto.fi

Abstract

Background: Chronic kidney disease (CKD) is a common condition worldwide and home dialysis (HD) provides economic, quality of life, and clinical advantages compared to other dialysis modalities. Human-centered design aims to support the development of eHealth solutions with high usability and user experience. However, research on the eHealth needs of patients using HD is scarce.

Objective: This study aimed to support the design of eHealth for patients with CKD, particularly for patients using HD, by developing a kidney disease experience map that illustrates user needs, concerns, and barriers. The research questions were (1) what experiences do patients, particularly older adults, have in their everyday lives with CKD? (2) what user needs do patients with CKD have for HD eHealth? (3) how can these needs be illustrated using the experience map technique? The study focused on patients aged >60 years, as they are at a higher risk of chronic conditions. The study was conducted as part of the eHealth in HD project, coordinated by Hospital District of Helsinki and Uusimaa, Finland.

Methods: In total, 18 patients in different care modalities participated in retrospective interviews conducted between October 2020 and April 2021. The interviews included a preliminary task with patient journey illustrations and questions about their experiences and everyday lives with CKD. The data analysis was conducted using a thematic analysis approach and the process included several phases.

Results: On the basis of the thematic analysis, 5 categories were identified: healthy habits, concerns about and barriers to eHealth use, digital communication, patients' emotions, and everyday life with CKD. These were illustrated in the first version of the kidney disease experience map. The patients had different healthy habits regarding social life, sports, and other activities. They had challenges with poorly functioning eHealth software and experienced other factors, such as a lack of interest and lack of skills for eHealth use. Technical devices do not always meet the emotional or physical needs of their users. This caused feelings of frustration, worry, and fear in patients, yet also fostered situational awareness and hope.

Conclusions: The experience map is a promising method for illustrating user needs and communicating the patient's voice for eHealth development. eHealth offers possibilities to support patient's everyday life with chronic disease. The patient's situation and capacity to use eHealth solutions vary with their everyday challenges, opportunities, and their current stage of treatment. The kidney disease experience map will be used and further developed in the ongoing research project "Better Health at Home—Optimized Human-Centered Care of Predialysis and Home Dialysis Patients" (2022 to 2026).

(JMIR Hum Factors 2025;12:e48221) doi:10.2196/48221

KEYWORDS

user need; chronic illness; kidney disease; older adult; eHealth; experience map; human-centered design; home dialysis

Introduction

Background

Chronic kidney disease (CKD) is a common condition worldwide. It has been estimated that 11% to 13% of the population in high-income countries have CKD. The number of patients with advanced kidney disease is growing 5% to 7% per year [1,2]. The prevalence of CKD is highest among older adults, ranging from 38% to 44% in patients aged >65 years [3,4]. In total 2.05 million people were treated with dialysis worldwide in 2010 [2]. In 2017, 1.2 million people died globally because of CKD [5].

Patients with end-stage CKD need kidney transplants or dialysis to survive. Dialysis can be performed in a dialysis unit in a hospital (ie, in-center dialysis), in a satellite dialysis unit, or at home using peritoneal dialysis or hemodialysis. The dialysis modality may vary depending on the patient's current health and life situation. Both in-center and satellite dialysis can be laborious for patients [6] and impose a heavy financial burden on medical care [7]. Therefore, home dialysis (HD) provides better quality of life and clinical advantages and empowers patients by providing them with more flexibility in their everyday lives [6,8].

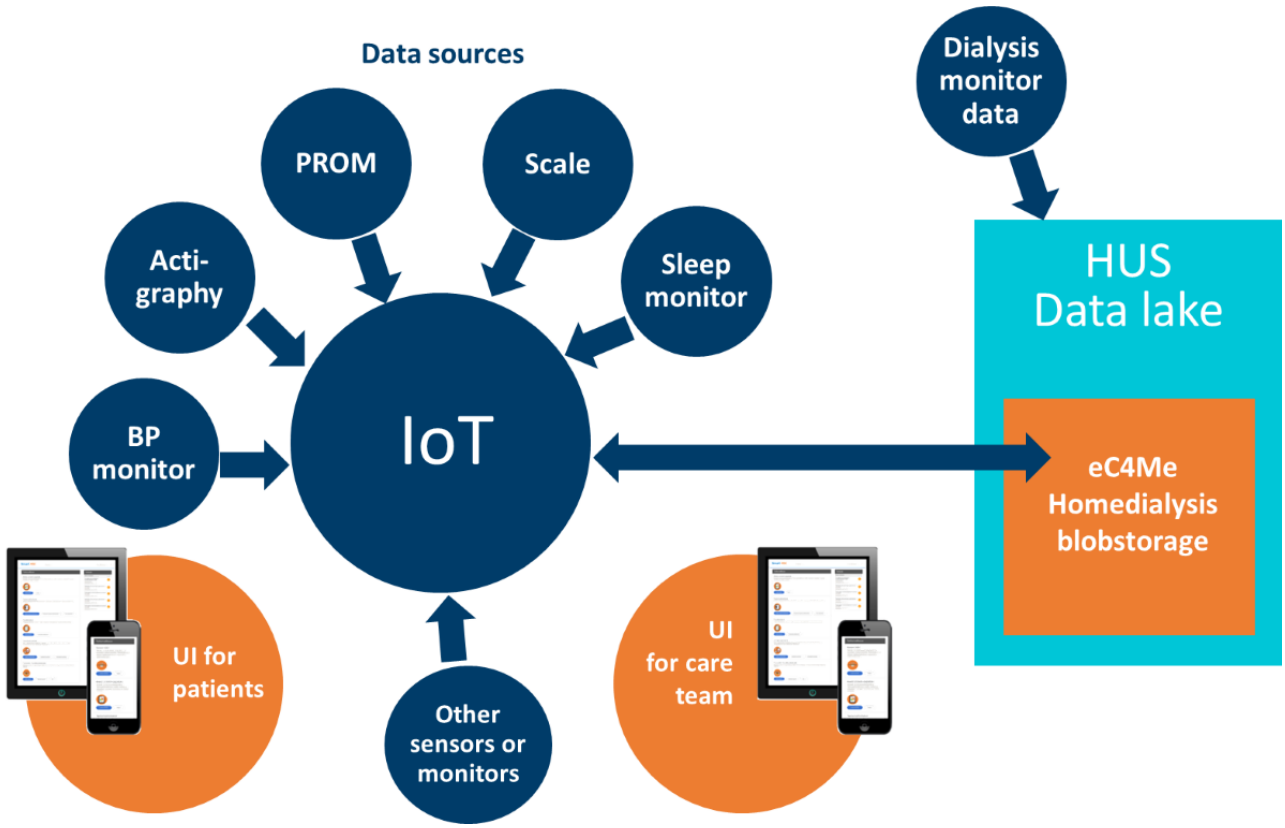
Even though bringing dialysis treatment to a patient's home might be burdensome and complicated for both the patient and the health care unit [8,9], there is a common understanding that HD prevalence needs to increase [1]. HD creates opportunities to improve the patient's safety and quality of life, as well as to support self-management of health [6,10]. HD has also proven to be feasible for older adults, even though challenges such as fear of needles or doubts about handling them, or other physical limitations have been recognized [8,11,12].

Research advises a holistic eHealth design that integrates technologies, end users, and use contexts [13-15]. Using human-centered design (HCD) approach [16], our study aimed to support the design of eHealth for patients with CKD, especially those on HD care, by investigating user needs, concerns, and digital service barriers, and illustrating those in the format of a kidney disease experience map. Due to the high prevalence of CKD among older adults [8], both in Finland and worldwide [17], the focus of the study was set on the perspective of older adults. The research questions were as follows:

1. What experiences do patients, particularly older adults, have in their everyday lives with CKD?
2. What user needs do patients with CKD have for HD eHealth?
3. How can these needs be illustrated using the experience map technique?

The study was part of the eHealth in HD project ("device research") [18] coordinated by the Hospital District of Helsinki and Uusimaa in Finland from 2020 to 2022. The project aimed to create a novel eHealth solution for patients with CKD undergoing HD (named "Home dialysis eHealth solution") and research devices that support patients both before and throughout dialysis therapy by gathering monitoring data, managing HD supply orders, and facilitating communication between the patient and the health care team (Figure 1).

Figure 1. eHealth solution and research devices for home dialysis. BP: blood pressure; HUS: Helsinki University Hospital; IoT: Internet of Things; PROM: patient-reported outcome measure; UI: user interface.



HCD of eHealth Solutions

HCD is an approach to interactive system development that aims to develop solutions with high usability and user experience [16]. The principles of HCD highlight the importance of involving end users in an iterative design process [16,19] and understanding their needs at the beginning of the development process [16,19].

When exploring user needs, researchers collect information, for example, about what features digital services should contain and how they should function from the perspectives of end users. The user needs exploration is important as it helps service designers to understand the role and meaning of the solutions for end users, in our case, patients using HD. eHealth is used to refer to digital health services, and health information offered through the internet or other technology solutions improving health care [20,21]. By designing in a human-centric way and involving end users in the development, eHealth can offer more value and new ways to provide health care and well-being services for patients and all citizens [13,21]. For patients with chronic illnesses, eHealth solutions can offer new communication possibilities between the patient and the health care team, as well as support empowerment and improve quality of life [21,22].

Patient Journey Maps and Experience Maps as Illustrations of User Needs

A patient journey map and an experience map are tools for 2 different purposes. In this study, patient journey maps were used as a part of patient interviews [23], and an experience map was used to communicate the results of the study holistically.

Patient journey maps as tools for visualizing patient journeys can be used to support the design of eHealth in several ways. The maps can help researchers and designers to identify how health care processes can be improved, for example, identify gaps and improve the processes by integrating eHealth solutions as part of those processes [24]. The patient journey maps sum up patients' experiences and care activities in a single, chronological, timeline-type visualization [24]. As the maps can help to explain what patients go through with their disease [23,24], they can help designers promote empathy, which is important in HCD [25-27]. Furthermore, they can help the sharing of knowledge between stakeholders as well as activate creative thinking [28].

Patient journey maps can be applied in various specialties regardless of the medical condition. For example, patient journey maps have been used to illustrate home isolation experiences of people with mild COVID-19 [29], hypertensive disorders in pregnancy [30], cancer [24], cervical dystonia [31], and kidney diseases [32]. In literature, the focus of patient journey maps varies from communicating patients' emotions and identifying process gaps to the importance of shared decision-making [33].

Experience maps are tools to combine situations, functions, emotions, and contacts in the same visualization [34]. Experience maps are valuable in the development of various eHealth, mobile health, and other apps, helping to communicate, for example, how cancer affects a patient's life and capturing patients' voices [35]. However, similar to patient journey maps, there is no standardized way to visualize and use them in an eHealth development context [33]. To our knowledge, no experience maps have been produced to represent the everyday life of patients with CKD using HD. Even though working with experience maps helps to identify users' needs [34], based on our literature review, few patient journey maps or experience maps focus on illustrating and communicating the needs as a basis for eHealth design and requirements specification.

Older Adults as eHealth Users

The European population is aging [36]. In Finland, >2 million citizens (ie, 36% of the population) are ≥55 years old [37]. Finnish older adults have good self-confidence in using digital services. In the 55 to 64 age group, >80% of Finns felt that their digital skills were at least on par with those of other Finns, and in the 65 to 74 age group, >60% of Finns felt the same [38]. At the same time, the number of patients in Finland who undergo dialysis, is expected to increase by >36% before 2040 [39]. Therefore, Finnish older adults seem to have both a need for and self-confidence in using CKD eHealth.

Older adults, due to an increase in diseases and complex health issues, could benefit substantially from eHealth [40]. Still, they experience several barriers to using eHealth, such as privacy concerns, lack of motivation to use digital solutions [41,42], and challenges to finding, accessing and understanding health-related information [40,43]. In addition, they have concerns about eHealth reducing the time with the physicians during appointments [42]. Although they can benefit from eHealth, they still face several challenges that could be solved with careful design.

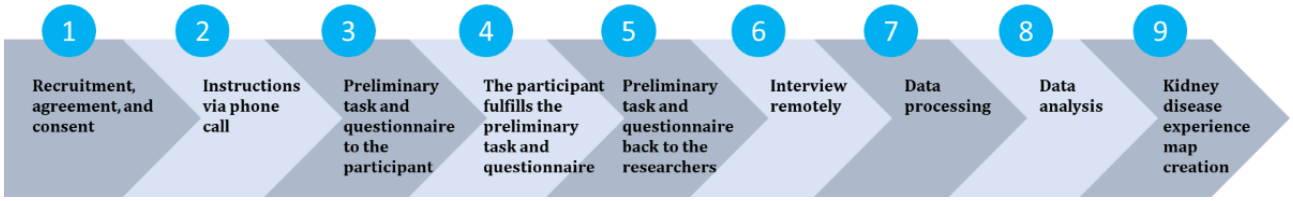
The health situation also affects the use of the eHealth and user needs. Many eHealth tools are intended for the treatment of a specific disease. One or more chronic health conditions can lead to several regular self-management tasks [40], which can potentially affect a patient's everyday life and the use context of eHealth. Therefore, eHealth should be adapted to each user's special situation [44].

Methods

Overview

Combining different context-specific strategies is important when developing services in a complex environment, such as health care [15]. Our study on user needs of patients with CKD for eHealth used a qualitative research approach and semistructured interviews [23,45] as the primary method for data gathering. The interviews were supported with preliminary tasks, which included visualizing patient's journey in the form of a timeline drawing [23]. The procedure of the study included several phases starting from the recruitment and ending with the creation of the kidney disease experience map illustration (Figure 2).

Figure 2. Procedure of the study.



Ethical Considerations

The empirical study was conducted as part of the eHealth in HD project, which received permission from the ethical committee of the Hospital District of Helsinki and Uusimaa (HUS/1649/2020, Jarkko Ihalainen). All participants gave their voluntary, informed, and written consent. The patients' capabilities to participate in the study were ensured. While analyzing the data, the information of the patients participating in the study was pseudonymized and coded with the identification numbers. Only the researchers assigned to the study had access to the data. No compensation was paid to the participants for their participation.

Participants

In total, 18 patients aged ≥60 years participated in the study: 5 patients in predialysis phase, 4 patients in satellite dialysis, 5 patients in home peritoneal dialysis care, and 4 patients in home hemodialysis care (Table 1). Most of the participants were retired. The sample excluded in-center patients with dialysis, but included patients with dialysis in satellite units. The participants were recruited from the group of patients who were participating in the larger research project of HD eHealth solution development.

Table 1. Demographics of study participants (N=18).

Demographics	Frequency, n (%)
Dialysis	
Predialysis	5 (28)
Satellite dialysis	4 (22)
Home peritoneal dialysis	5 (28)
Home hemodialysis	4 (22)
Gender	
Woman	4 (22)
Man	14 (78)
Education	
Basic education	4 (22)
Upper secondary level	4 (22)
Bachelor's degree	5 (28)
Master's degree	3 (17)
Other	2 (11)
Technology skills (self-assessed)	
Good	6 (33)
Basic	11 (61)
Weak	1 (6)

The health care team recruited the patients by distributing materials about the study and asking about their interest in participation. Participation was voluntary and they did not get any compensation. The study participants were already familiar with the novel HD eHealth solution under development and had tested the first version of the solution. The research nurse contacted the potential participants first and informed them about the study. If the participants wanted to participate in the study, they signed the consent form, and the research nurse provided their contact information to the researchers. Then, the research material package, prepared by the researchers, consisting of a cover letter, a preliminary interview task, background information forms, and the responsible researcher's contact details were sent to the participants. After sending the package to the participants, the researcher called them to provide more detailed instructions, answer any questions, and schedule a time for the remote interview.

Preliminary Tasks and Retrospective Interviews

Interviews are useful in the phases of any development process related to eHealth [15]. In our study, retrospective interviews included questions about the patient's journey, treatments experiences, cooperation with the health care team, technology experiences, and visions for the future. The themes broadly covered patients' everyday experiences with the illness, practical questions, and the comprehensive timeline with CKD.

The idea of the visual timeline drawing as the preliminary task for the interview was to (1) help the interviewee to process and structure their multistage patient path even before the interview, and (2) help with communication between the interviewee and researcher during the remote interview [23]. The preliminary task was sent to the interviewees 2 weeks before each interview. In this task, patients were asked to identify and illustrate significant milestones, events, and experiences with their illness to a timeline. The patients returned the task to the researchers before the remote interview.

The interviews were conducted between October 2020 and April 2021 when the second wave of COVID-19 was underway in Finland. For safety reasons, the interviews were arranged remotely via Microsoft Teams. Two researchers were present in the interview session with the participating patient: one researcher being the interviewer and the other note-taker. Interviews were audio-recorded using recording functionality of Microsoft Teams. Data gathering was conducted in collaboration with 5 researchers (SL, NK, JV, SH, and PV): 2 doctoral researchers, 2 students, and a professor from the human-computer interaction field.

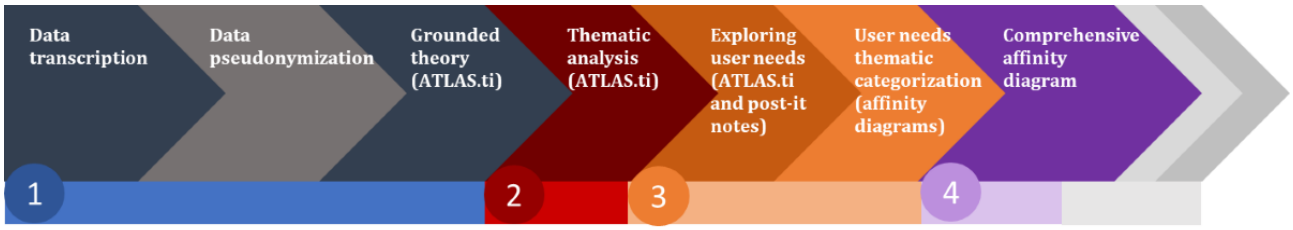
Data and Analysis

Overview

The data included recordings and notes from the interviews, as well as illustrations of patient journeys. The interview notes were finalized according to audio recordings and the transcriptions were pseudonymized. In total, the data consisted of 155 pages (ie, 58,429 words) of written notes and 18 paper-based patient journey visualizations (ie, visual timeline drawings). A total of 5 researchers (SL, NK, JV, SH, and PV) participated in the data analysis.

The data were analyzed in 4 phases (Figure 3): (1) grounded theory analysis, (2) thematic analysis, (3) user needs exploration and thematic categorization, and (4) comprehensive affinity diagram creation. The data included many different perspectives to understand the different nuances and ensure reliability, and the analysis was done in collaboration with several researchers in many phases.

Figure 3. Four phases of data analysis.



Phase 1: Grounded Theory Analysis

In the first phase [46], the pseudonymized interview transcripts were analyzed following the main structure of grounded theory [47,48] using ATLAS.ti software version 9.1.5.0 (ATLAS.ti Scientific Software Development GmbH, "ATLAS.ti") [46]. During the analysis, the interview data was coded using researcher-denoted concepts and open coding influenced by the grounded theory [46,47]. The first round of interview data analysis following grounded theory was conducted by SL.

Phase 2: Thematic Analysis

In the second phase, the principles of the thematic analysis method [48,49] were followed to analyze the interview data from 18 participants from an existentialist perspective, including references to death, well-being, and atmosphere in life (eg, emotions) [50-52]. In addition, the analysis used the holistic framework [13-15] of exploring the context of use, technology, and people.

On the basis of the analysis, 3 thematic categories were identified: healthy habits, concerns and barriers for eHealth use, and everyday life with CKD. Observations based on interview transcripts were written on post-it notes and thematically categorized following the phases of the affinity diagram method [48,49]. The data gathered using preliminary task of patient journeys were mapped in line with the interview transcripts. In addition to the initial 3 themes, 2 categories were formed based on thematic analysis: digital communication and patients' emotions. Concerns and use barriers for eHealth were investigated from the data by two researchers (PV and SH). The digital communication affinity diagram was done by three researchers (PV, SH, and JV).

The thematic analysis categories include several aspects as presented in Textbox 1.

Textbox 1. Categories of thematic analysis.

- Healthy habits: health-promoting activities and hobbies that were mentioned by interviewees, including the whole scale of activities from sports to eating habits.
- Concerns and use barriers for eHealth: observations related to everyday life challenges with the disease, use challenges and barriers of digital services, negative emotions, such as worries or dissatisfactions, and patient path challenges.
- Digital communication: observations concerning health care-related digital communication habits, channels, and experiences.
- Patients' emotions: observations of emotions and experiences interviewees mentioned in the interviews and patient journey visualizations. The emotions included the whole scale of emotions, from fear and confusion to happiness.
- Everyday life with chronic kidney disease: observations related to interviewees' thoughts, stories, and experiences of their everyday lives, including references to death. Death made the participants reflect from different perspectives: their own (forthcoming) death was compared to other deaths, the actions to avoid death were listed, and the inevitable nature of death was considered.

Finally, the frequency of observations per theme was calculated (Multimedia Appendix 1). Most of the observations were related to concerns and barriers to technology use. Emotions and healthy habits were also common topics in all the interviews.

Phase 3: The User Needs Exploration and Thematic Categorization

The third phase focused on analyzing user needs. All the gathered data were explored from the viewpoint of user needs from 2 perspectives: exploring digital communication needs and exploring overall user needs for interactive solution development. In the analysis, 2 researchers (PV and SH) explored the data labeled "digital communication," "user needs," and "to eHealth solution concept" in ATLAS.ti and formulated the observations as user needs. The second round resulted in 165 needs.

After this, the digital communication affinity diagram was created in collaboration with 3 researchers (SH, JV, and PV). The user needs were written on post-it-notes by PV, and analyzed, and regrouped to the affinity diagram in collaboration with SH, JV, and PV. In the affinity diagram, the identified needs (n=165, 100%) were thematically grouped under 5 categories: overall interaction and communication (n=77, 47% of all identified needs), patients' digital activities (n=28, 17%), instruments (n=27, 16%), inventory and ordering dialysis supplies (n=19, 12%), and digital communication with the health care team (n=14, 8%). While creating the affinity diagram of digital communication needs, doubles were removed.

When exploring overall user needs, the data transcriptions of 18 interviews were analyzed using ATLAS.ti. During this analysis, the data labeled with the following codes—"user needs," "healthy habits and hobbies," "suggestion," "communication," "digital services," "tasks," "care team," "communication," "social relationships," "equipment," "challenges," "positive experiences," and "device research"—were explored and formulated as user needs. In total, 287 user needs were identified and written on post-it notes as preparation for the comprehensive affinity diagram work.

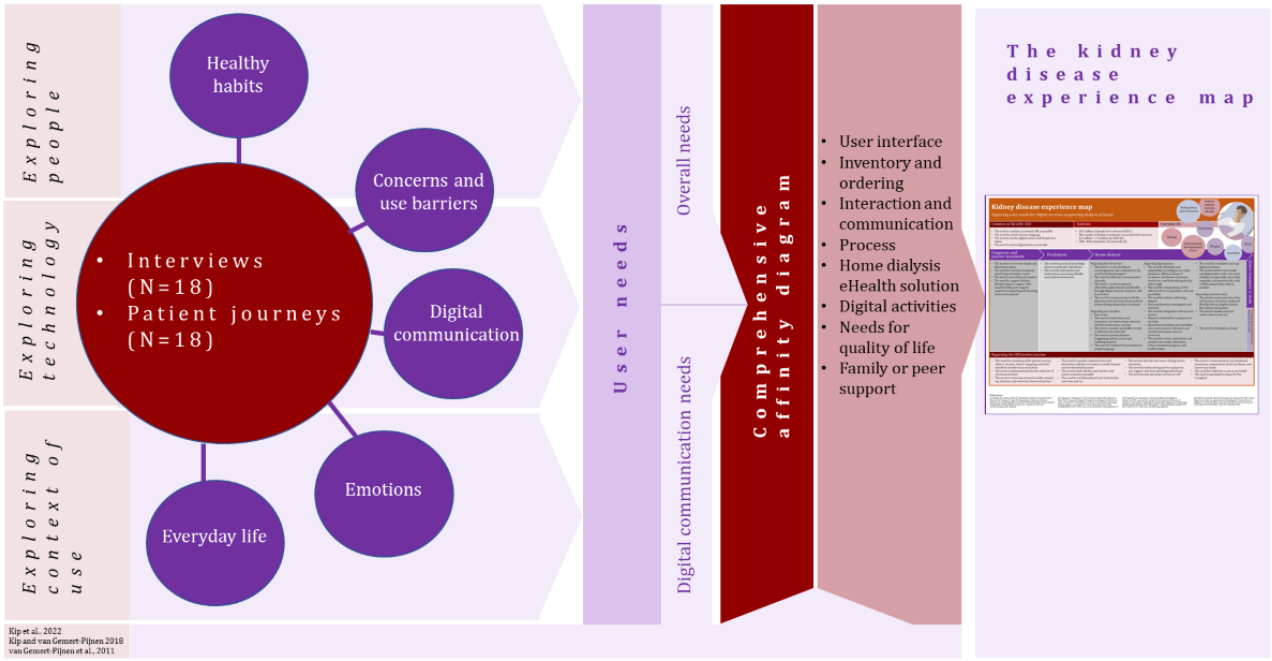
Phase 4: Comprehensive Affinity Diagram

In the fourth phase, the affinity diagrams created in the previous analysis phase were merged to form a comprehensive affinity diagram. Three researchers recategorized the post-it notes to create a comprehensive affinity diagram of user needs. The main themes of the diagram were "eHealth user interfaces," "Inventory and ordering dialysis supplies," "Overall interaction and communication," "Process," "HD eHealth solution," "Family/peer-support," "Patients' digital activities," and "Needs for improving the quality of life" (Multimedia Appendix 2). These 5 were used as the leading themes to guide the design of the first version of the kidney disease experience map. Data analysis results were discussed between PV, SH, VR, and JV.

Creation of the First Version of the Kidney Disease Experience Map

An experience map is a tool that strings the perspectives of the context of use, people, and technology, together in a holistic manner [13-15]. The maps can be used to support the communication between stakeholders and to capture a patient's voice in eHealth design [35]. The first version of the kidney disease patient experience map (Figure 4) was made by combining remarks from existing example models particularly the cancer experience map [35]. In the first paper-based template of our map, the preliminary phases from the preliminary tasks (ie, patient journey maps) were illustrated: "before diagnosis," "diagnosis and passive treatment," "treatment method selection," "beginning of treatment/training," and "home dialysis." Post-it notes from the comprehensive affinity diagram were added under those phases. The first draft of the patient experience map was created in collaboration with 3 researchers.

Figure 4. Overview of the data analysis process leading to the creation of the first version of the kidney disease experience map.



Next, the structure of the experience map was refined based on the number of post-it notes in each phase. The duplicate post-it notes were removed and the content of the map was reorganized (PV, SH, and JV). To validate the contents of the draft map further, 2 researchers (PV and SH) reexamined the post-it notes after which 4 other researchers (JV, NK, SL, and VR) gave feedback on the next version of the map. After all, in total of 6 researchers (PV, SH, JV, NK, SL, and VR) participated at least somehow in the experience map visualization process. The published version of the first version of the kidney disease experience map was finalized by PV. An overview of the multiphase analysis process which led to the creation of this first version of the kidney disease experience map is illustrated in Figure 4.

Results

Overview

As the outcome of the study, the first version of the kidney disease experience map, which illustrates the user needs, concerns, and barriers of patients with CKD for eHealth, was created. This section explores the patients (ie, people), technology, and context of use perspectives [13-15]. First, patients' healthy habits are investigated and their concerns and barriers to technology are addressed. Then, the technology is examined, concentrating on digital communication. Finally, the context of use from the perspective of patients' everyday lives is explored. Due to sensitive personal issues, research results and citations are presented without identification information to ensure complete anonymity of the participants.

Exploring the People: Healthy Habits

The patients had different needs for healthy habits regarding social life, sports, and other activities (Multimedia Appendix 3). Patients mentioned that social life and taking care of close relatives, such as grandchildren, were important to them. Incidental exercises, including gardening or shoveling snow, were popular. Being in nature, picking mushrooms or berries, fishing, and hiking, along with other everyday life activities, such as club and society activities or studying, brought well-being to the participants. Animals, for example, dogs, cats, and summer chickens, also supported the participants' well-being. Handcrafting, art, and culture were also seen as important, but COVID-19 restricted those activities. The importance of many remote social contacts was mentioned in the data, and patients had to learn new eating habits, such as losing weight or cutting down on alcohol consumption.

Exploring the People: Concerns and Barriers for eHealth Use

Patients were challenged with poorly functioning available eHealth solutions and also experienced other negative factors, such as lack of interest or lack of skills (Multimedia Appendix 4). Technical devices did not always meet the emotional or physical needs of their users. This caused emotions like frustration, worry, and fear from patients. The patients' responses covered their available eHealth solutions, not focusing especially on the HD eHealth solution.

The solution or software used by patients for their CKD care was found to function poorly in many ways. For example, the devices as a part of the HD eHealth solution did not work as expected, or they worked differently than they expected:

> *I have no control over the information. If I measure my blood pressure, I press the BT button and then it should be delivered, but I don't know if it did. The same goes for the scale, *beep*, and maybe it synchronized.* [P18]

Alongside poorly functioning devices, in some cases, the HD eHealth app also did not work well. The software content (eg, one's results or visualizations of one's health situations) was challenging to understand or find. The HD eHealth app was challenging to use with small mobile phone screens and its user interface terminology caused problems for patients:

> *There have been a few problems with the phone. There will be updates and then it will freeze somehow. I don't understand much; the basic things are in Finnish, but then they are English words so... This was at least the third time that the nurse had visited us.* [P17]

> *Others have also worked well. Even the mobile phone. You can monitor your health from it, even though it has those [health status] curves that you don't understand much about.* [P13]

The participants doubted the functionality of the HD eHealth solution, both devices and the app. With the dialysis monitor, they were worried that something would go wrong:

> *I haven't used (ie, the devices and instruments) much, I've just tried. I'm a bit lazy. I doubt whether the instrument is still in good condition. There was an error message about a month ago. There was a notification that there was still something wrong. I thought I'd let it be and fix the instrument before I started.* [P24]

> *When the first treatments started, it was so exciting. But after a month had passed, it had become routine and was no longer exciting. Then, when I moved home from hospital, I started to get excited again.* [P22]

Exploring the Technology: Digital Communication

Overview

Patients commented on digital devices and the app from many different aspects in terms of the technical functionality and the emotion they cause. The contents of the solution's user interface (Figure 1), as well as other user interfaces, were also commented on. Patients linked and communicated with the health care team through digital channels and used technology independently as a part of their HD treatment during the research.

Many Groups Communicate With the Patient

Communication within the health care team is important, but also nurses, physicians, and patients share information (Multimedia Appendix 2). To get the supplies and instruments needed for HD care, the patient also needs to be in regular contact with the pharmacy or the health care team. Relatives must not be forgotten either, as patients may want to share information with them:

> *To be able to get in touch with the on-call nephrologist or the other way around, perhaps you could organize a little more frequent meeting. I think that many patients could have questions for the doctor that the nurses cannot or are not allowed to answer. Nephrologist on-call from the computer.* [P24]

Examples of Current Communication Channels

Depending on the patients' devices and patients' capabilities to use them, HD patients use different communication channels for following their health records and CKD-specific operations. They order HD supplies once per month. Some patients have created their digital tools (eg, Excel sheets), which calculate the number of needed supplies automatically. After that, the order will be sent via email to the nurse. Still, Excel sheets or other tools for ordering dialysis supplies are not an option for everybody. Therefore, an easy-to-use solution for ordering supplies and managing inventory is needed for HD patients:

> *Orders must be sent once a month. I have an Excel sheet that automatically keeps the balances. Then, there is a page where you can automatically see what needs to be ordered. I have made Excel sheets myself. It's quite a job when there are more than 30 different supplies. Syringes... you must know their numbers. Sometimes, I also check some information. If one item is true, then so are all the others. It's automatic. I then make a .pdf list from the Excel sheet, one A4, and then it goes [by e-mail to the nurse].* [P30]

In Finland, patients follow their health records via national or local patient portals (eg, "My Kanta Pages" [53] or "MyChart" [54]). Patients with chronic illness have many laboratory tests taken. They appreciate that all their health records are found in one place, but each record is shown in detail separately, too. Therefore, easy-to-read visualizations of test results and current health situations are needed:

> *MyChart is as hopeless as My Kanta, both of which break down the lab results separately, even if they want to see them as a whole. Applying for your results is difficult.* [P37]

> *I look at the results on 'My Kanta,' all diagnoses come there, as well as laboratory test results. From there, I can see them all. The information is readable and remains stored. I think 'My Kanta' is really good. There you can also see laboratory visits from a certain period if you want to follow a limited period.* [P21]

Positive Attitude Toward Digital Communication

All patients participating in the study seemed to have a positive attitude toward digital communication tools. For example, paper-based dialysis supply orderings may be forgotten and instead, digital tools can step in and help patients with their everyday lives. Patients felt that digital solutions enabled information-seeking and offered more flexibility in meetings with the health care team and other contacts:

> *Their version is that they hand out an A4 paper. You should fill it in; it's difficult manually. The risk is that you forget.* [P30]

> *We have internet and e-mail. My wife uses them more; I just read. The wife handles banking and other things remotely. I kind of stopped using the computer when I don't need to. You could probably handle things digitally. I haven't come across anything like that. I have had to deal with a lot of paper. I have not been approached digitally. Yes, it would be even easier. You could reach the staff at any time.* [P24]

Exploring the Context of Use: Everyday Life and Emotions

Overview

The main themes related to everyday life were thinking of death, the importance of care, changes in well-being after treatment started, and emotions (Multimedia Appendix 5).

The Presence of Death

With kidney disease, death is present in life. It is known that without dialysis or receiving a transplant, one dies. The logic of the disease is unequivocal: when dialysis no longer helps, the person knows they are going to die:

> *I think that the disease is deadly and must be treated.* [P21]
>
> *Everyone knows that life ends eventually, but I'm not afraid of it in any way.* [P1]

The General Well-Being of the Participants Varied

The patients understood the essentiality of treatment and often felt better after the treatment started. They wished to be able to continue life as normally as possible. One patient (P9) described experiencing leg cramps at night, which had caused them to sleep poorly. After the dialysis treatment started, the cramps stopped. They managed better than before and could even go for walks:

> *I hope I don't end up on dialysis. Of course, it is ahead, but hopefully as far as possible. It is a wish.* [P20]
>
> *It's just that now I feel like I'm getting older. I recently turned 70, so all kinds of unnecessary aches and pains will increase.* [P25]

Everyday Life Is Characterized by Waiting

The health situation of a patient can be improved either with dialysis or a transplant. Having to wait for transplant can have major setbacks; one patient (P24) had to be removed from the transplantation waiting list due to amputation and sepsis, which is why they now focus on keeping themselves in the best possible physical condition through physical activities. Patients were waiting to travel.

A Wide Range of Emotions in Everyday Life

A wide range of emotions appeared in the interview data about the patient's everyday life. Despite the challenges of the disease, most of the patients described their lives with the disease in somewhat positive terms. Everyday life with the disease seemed to go somewhat smoothly.

In addition to positive feelings, patients also expressed negative emotions, such as confusion, fear, irritation, anger, and frustration. Feelings of confusion were caused by changes in health status and events in the patient's journey, uncertainty of the future steps in the journey, or technology that worked unexpectedly. After receiving the diagnosis, one patient was confused because they did not fully understand what kind of disease the diagnosis was about. Several patients brought up fear in the interviews. Things related to the treatments were especially scary. Most of the patients also described irritation or anger at least sometimes during the patient journey. In addition, half of the patients reported frustration. Health challenges, the difficulty or even failure of a treatment, and the commitment to dialysis treatment were irritating. The most frustrating thing was when HD treatment did not go as well as planned. Another source of frustration was digital systems, equipment, and devices:

> *And a little fear. I am quite hopeful that I will get there on the transplantation waiting list and that it will be successful at some point. I hope that I can make it there (ie, from a health perspective), and if this continues, then there is nothing to worry about.* [P11]
>
> *Actigraphy, a part of HD eHealth solution, is inconvenient, but there is nothing wrong with the other devices, everything is fine. I have a strong belief that information will go forward. It's enough for me to see on the scale and blood pressure monitor that my weight and blood pressure are okay.* [P18]
>
> *This home dialysis takes up a lot of space. Do you see these boxes? A huge package comes once a month. The dialysis monitor also takes up quite a lot of space. And then the cabinets, where you put all the treatment consumables. And there is a bottle like this: hand sanitizer comes once a month. Quite a lot of supplies come from the hospital once a month. This room is completely reserved for dialysis.* [P30]

HD taking up too much space or other resources was both irritating and frustrating. HD requires much space at home, so some patients had to renovate their homes. However, almost half of the patients felt that their living conditions did not require changes due to HD. The patients felt it was important that the dialysis treatment could also be done at the summer cottage.

Positive emotions included joy, optimism, hope, and satisfaction. A third of the patients mentioned joy in the interviews and most patients expressed hope and optimism. They were satisfied with the professional staff and the stability of the current situation. The supportive and successful dialysis made them happy. Keeping their health situation stable was their biggest wish. Patients also mentioned their willingness to continue their hobbies.

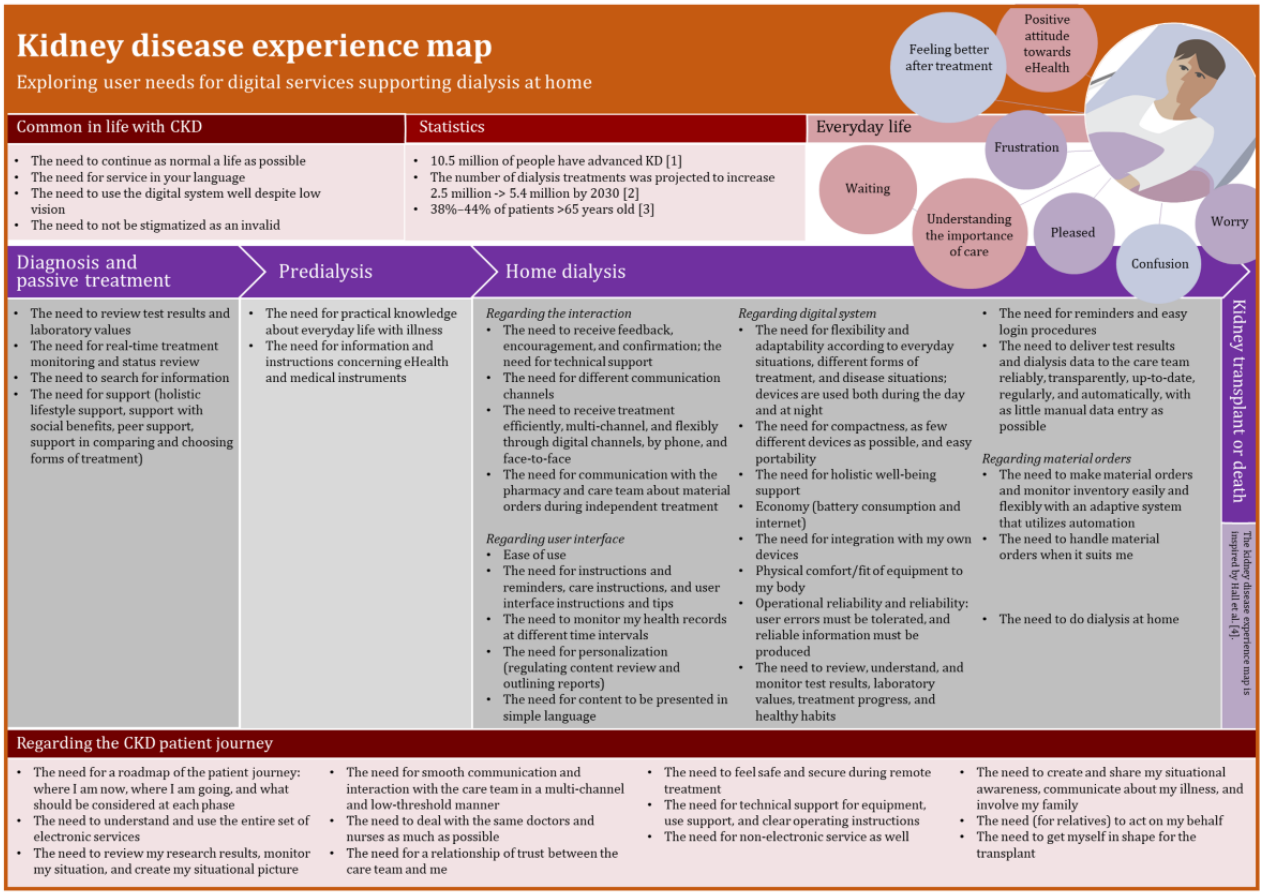
The patients were satisfied with both their health condition and the information related to it. Half of them were satisfied with the treatment. The patients were pleased with the eHealth solution in use during the study and the HD treatment was experienced as proceeding smoothly. The nursing staff was considered friendly and knowledgeable:

> *I have been very satisfied. They, the care team, have guided me, and when I have challenged them, they have accepted challenges and promised to sort things out.* [P37]
>
> *Home dialysis has gone surprisingly well—no major problems.* [P34]
>
> *At least now, this present moment looks bright.* [P13]

The First Version of the Kidney Disease Experience Map

The first version of the kidney disease experience map (Figure 5 [1-3,35]) was developed from the perspective of user needs of eHealth. It includes their needs for everyday life with CKD (ie, "Common needs"), needs relating to 4 phases of the journey of patients with CKD, and needs in all journey phases of patient with CKD from diagnosis to kidney transplant or death. The patient journey phases, which are inspired by the experience timeline of cancer experience map [35], are "Diagnosis and passive treatment," "Predialysis," "HD," and "Preparing for kidney transplant or death." In the top corner of our map kidney disease statistics and patients' everyday life experiences with CKD are illustrated.

Figure 5. The first version of the kidney disease experience map from a user needs perspective. A higher resolution image is available in Multimedia Appendix 6. CKD: chronic kidney disease. KD: kidney disease.



References:
[1] Mendu ML, Divino-Filho JC, Vanholder R, Mitra S, Davies SJ, Jha V, Damster KC, Gallego D, Seger M. Expanding Utilization of Home Dialysis: An Action Agenda From the First International Home Dialysis Roundtable. Kidney Medicine 2021 Jul 1;3(4):635–643. doi: 10.1016/j.xkme.2021.04.004
[2] Liyanage T, Ninomiya T, Jha V, Neal B, Patrice HM, Okpechi I, Zhao M, Lv J, Garg AX, Knight J, Rodgers A, Gallagher M, Kotwal S, Cass A, Perkovic V. Worldwide access to treatment for end-stage kidney disease: a systematic review. The Lancet 2015 May 16;385(9981):1975–1982. doi: 10.1016/S0140-6736(14)61601-9
[3] Vanholder R, Annemans L, Bello AK, Bikbov B, Gallego D, Gansevoort RT, Lameire N, Luyckx VA, Noruisiene E, Oostrom T, Wanner C, Wieringa F. Fighting the unbearable lightness of neglecting kidney health: the decade of the kidney*. Clinical Kidney Journal 2021 Jul 1;14(7):1719–1730. doi: 10.1093/ckj/sfab070
[4] Hall LK, Kunz BF, Davis EV, Dawson RJ, Powers RS. The Cancer Experience Map: An Approach to Including the Patient Voice in Supportive Care Solutions. Journal of Medical Internet Research 2015 May 28;17(5):e3652. doi: 10.2196/jmir.3652

The first version of the kidney disease experience map (Figure 5) includes aspects from all three perspectives [13-15]:

1. Context of use: the context of use perspective can be found especially in the patient journey phases and the "everyday life with CKD" part of the map.
2. People: information about people is available in the "statistics and everyday life with CKD" parts of the map.
3. Technology: technology aspects are mainly covered in the form of user needs.

Discussion

Principal Findings

On the basis of our study, eHealth as a part of the daily lives of the patients with CKD was perceived as being useful, but the patients also had challenges with devices and apps. Acknowledging that CKD is a long-lasting journey, it is essential to consider long-term user experiences. Finally, several user needs in different phases of the journey of patients with CKD were identified and illustrated with the first version of the kidney disease experience map.

Answers to Research Questions

Overview

This study had three research questions: (1) what experiences do patients, particularly older adults, have in their everyday lives with CKD; (2) what user needs do patients with CKD have for HD eHealth; and (3) how can these needs be illustrated using the experience map technique?

What Experiences Do Patients, Particularly Older Adults, Have in Their Lives With CKD?

Three key experiences related to daily life of patients with CKD were identified as depicted in Textbox 2

Textbox 2. Key experiences related to daily life of patients with chronic kidney disease (CKD).

- Patients' healthy habits: patients with CKD must adapt to the changes in their everyday lives resulting from dialysis care. For example, patients must learn new routines for their treatment at home. Patients felt that CKD limited their lives, but they dealt with it in different ways: some were frustrated and others continued to live as normally as possible. However, they understood the necessity of the treatment and continued the treatment as usual. Similar findings were reported from a study involving cancer patients [35], who expressed concerns with unknown, life-threatening, and life-changing experiences.
- Patients' concerns and barriers to use for eHealth: patients had challenges with user interface and technical devices, including dialysis monitors and research devices of HD eHealth solution (Figure 1), that did not always meet the users' needs.
- Digital communication: digital communication opportunities via the eHealth solution play an important role for patients. Digital solutions made information-seeking possible and offered more flexibility when meeting the health care team and other contacts. Information gaps did not accumulate as easily as for cancer [24].

What User Needs Do Patients With CKD Have for HD eHealth?

In our study, both general needs and needs related to different patient journey phases were identified and illustrated in the first version of the kidney disease experience map (Figure 5). We found that CKD was different from day to day, and thus the use context of eHealth varied. Poor usability burdens the everyday life of a person with disease even more. Similar observations were reported in the case of digital independent living supporting systems of older adults [34]: designers were important to understand the treatment stage, the transition from health to illness, and statuses between them. Furthermore, the needs and experiences of patients with cancer were found to vary during the care process [35].

The HD eHealth solution is targeted to all patients with CKD, as well as individuals with disabilities, such as low vision or even amputated limbs. The solution is to be used during the entire patient journey, which may last for years. Considering this, the HD eHealth solution is observed to play a remarkable role in the patient's life and highlights the need for research on long-term user experience.

How Can the User Needs Be Illustrated Using the Experience Map Technique?

Typically, user needs are illustrated with use cases or user needs tables, and they can be processed further to requirements format [55]. Findings from user research can also be presented as user profile illustrations, such as personas [15,34], stakeholder maps [15], experience maps [34], and visualizations of stakeholders [34].

In our study, we used the experience map method to illustrate user needs. As with the cancer experience map [35], we identified the phases of care and relevant themes for grouping the findings. We considered the experience map technique applicable for illustrating not only the phases but also the eHealth related needs of the patients. We found a group of general needs, which are common for all phases of care, but also needs that arise from specific phases of care. Understanding the care process and how user needs evolve is important for the designers and also for the health care team, who are responsible for introducing the services to patients and training them. Experience maps help to ensure the care continuum in service development [34].

In this study, the created experience map represents a common higher-level experience map, which can illustrate a specific persona, or the timeline could be narrower (ie, 1 month or 1 day). In the first version of the kidney disease experience map, the HD eHealth solution covers different wearable solutions and the software (ie, including the wearable solutions and software connected to the dialysis monitors); however, different physical dimensions (ie, dialysis monitors, packing materials of supplies ordered) are not covered.

Experience mapping can be seen as a promising but laborious method for illustrating user needs and communicating a patient's voice holistically to eHealth development, especially in a case where the experience map is based on extensive qualitative data. Following the example of cancer experience map [35], we captured patient's voices with their quotes. Similarly, we also focused on reporting patient quotes in the paper instead of the experience map. Therefore, the first version of the kidney disease experience map can be used both separately and with the paper. Our interpretation is that similar to the integrated patient journey map for hypertensive disorders in pregnancy [30] and the cancer experience map [35], the kidney disease experience map can help developers build effective eHealth solutions and promote empathy and shared understanding between the developing team members. It can also help empower patients and help with shared decision-making [33]. Like the patient journey map [24], the first version of the kidney disease experience map has the potential to help simplify and optimize processes for better health outcomes. In addition to that, similar to the cervical dystonia patient journey map [31], it can help to communicate the different phases of the care to the patients and help them to understand their needs at each phase.

Relevance of the Research

In the study, the experience map describing the patient's experience was formed based on the qualitative data from retrospective interviews and preliminary tasks. This first version of the kidney disease experience map can be used as a basis for other experience maps related to kidney diseases in the future. In addition, it can also bring perspective to the experience maps of other chronic diseases.

The presented results can be used to support the further development of HD eHealth. The results have already been used in HD eHealth solution development project in Finland (eg, to inform the redesign of the solution user interfaces).

Experience maps, such as the map created in the study, can help researchers, designers, and developers to empathize with the patients with chronic illnesses, and to understand their needs and use contexts for future eHealth innovations. We see our kidney disease experience map and other similar maps having the potential to help design the solutions from a more holistic perspective compared to currently prevailing practices [13,21], as well as to create a base for the new ecosystems and new ways to offer health care in the future [21]. We continue the work in the ongoing research project "Better Health at Home—Optimized Human-Centered Care of Predialysis and Home Dialysis Patients" (2022 to 2026).

Evaluation of the Study

Our study had some limitations. First, exploring the context of use was not allowed at patients' homes due to COVID-19 restrictions. Therefore, we needed to collect data remotely. This was partly problematic because the patient's home is the main context of using HD eHealth. In contrast, due to the patient journey preliminary task and participants' open-minded attitude during the interviews, we were able to gather rich data about user needs and the context of eHealth use. In our study, the possible impact of the COVID-19 period was considered.

Our sample size is quite small, however, typical for a qualitative interview study. The representativeness of our sample can be questioned as it is divided from a larger patient group based on age and treatment methods. To find as comprehensive a sample as possible, while ensuring the safety of the participants with their ongoing treatment, the health care team helped researchers in recruiting participants for the research study.

Due to the personal nature of health information and privacy, we could not release the participants' background information, such as specific ages or treatments in this paper. In addition, the indirect identifiers have been removed from the paper. The relative satisfaction of the patients in their lives despite CKD could have been influenced by various factors, besides the HD eHealth solution. First, the target group in this study included only active patients with relatively good physical conditions for research ethics reasons. Second, many of them were waiting for a kidney transplant, which brought them hope with the disease. Overall, there was a lot of variation in emotions, but deep gloom appeared in the data only rarely.

The first version of the kidney disease experience map was created based on a reasonable amount of qualitative data of the everyday lives on patients aged ≥60 years in Finland. Before participating in this study, the patients already knew that the study covered new eHealth solution testing. Therefore, perspectives from patients whose condition is very weak and those whose attitude is not open to new technology are missing in this study. In contrast, research shows that HD is optimal for those who live in areas with long distances to travel to an in-center dialysis unit [6,7]. In part, this can also be seen in the first version of the kidney disease experience map: the study did not include patient experiences during the kidney transplant phase—meaning that the user needs related to this essential phase of the patient journey are missing from our map. In future, the introduced version of the kidney disease experience map can be updated by including motivational issues for HD.

The study was conducted as an integrated part of the patient's normal CKD treatment, so the research setup was organized in line with the patients' typical daily conditions and treatments. The study included participants who started using the HD eHealth solution shortly before the interviews but also those who had already been using the solution for several months. Thus, the study covered both novice users and more experienced users. Patients' reported technology skills varied from weak to good.

Further Research

This study provided initial experiences of creating an experience map to illustrate patients' needs for eHealth. In addition, the patients' experiences from the HD eHealth solution development can offer important perspectives to other development projects regarding patients with CKD. Further research is needed to gather more experience about the applicability of the method. We aim to continue the research with patients with CKD and HD eHealth designs. We will use the illustrated first version of the kidney disease experience map and the listed user needs in further analysis, which will aim to describe design principles and guidelines for HCD of eHealth for patients using HD and other patients with chronic illnesses.

The outcome of the first version of the kidney disease experience map offers insights to the different phases of the patient journey for patients with kidney disease at the HD care. Due to the diverse experiences of patients in different phases of the care process, different types of kidney diseases could also be studied and visualized with experience maps. In addition, the next versions of the kidney disease experience maps could focus on specific days and months. The next versions of the kidney disease experience maps could also be created from the health care team's perspective.

The content of the first version of the kidney disease experience map was collected with qualitative research methods, which made the consideration of the many different nuances possible. Due to the qualitative nature of the data collection for the first version of the kidney disease experience map, we did not prioritize the available content, for example, based on the amount of the post-it notes. However, without careful prioritization, there is always a risk that the map is too challenging to use as a part of the practical software development work. Therefore, we will involve the patients, as well as the health care team, designers, and software developers, in prioritizing the content of the map in the future.

The perspectives of patients using in-center dialysis are missing from that data, which was the base of the first version of the kidney disease experience map. However, we believe the creation of experience maps for the eHealth needs of patients using in-center dialysis is important for the future as the dialysis modalities might change depending on the patient's health and life situation.

Our study pointed out the need for further research on the existentialism perspectives of the patients with chronic illnesses using eHealth. Deeper understanding is required on user needs and the role of eHealth as a part of patients' everyday lives, even when life is nearing its end [50-52].

Conclusions

eHealth offers possibilities to support the key needs of patients with chronic diseases, such as CKD; however, digital services need to adapt to a wide variety of situations and user needs because kidney diseases are unique to each patient. Patient experiences vary both between and within the stages of treatment. In addition, the patient's situation, needs, and capacity to use eHealth vary with everyday challenges and opportunities. Furthermore, with older adult patients, the problems that come with age increase the spectrum of barriers to use eHealth and associated challenges.

eHealth should enable proactive communication between the patient and the health care team. Furthermore, awareness of the status should be formed for both parties in real time, as much as possible. In this way, the patients feel safe, are more motivated in their treatment, and take responsibility for themselves.

The experience map is a promising method for illustrating user needs for eHealth, communicating a patient's voice and describing the longitudinal patient journey to support eHealth development. Our first version of the kidney disease experience map combines people, context of use, and technology perspectives holistically and illustrates the eHealth needs of patients with CKD for the future.

Acknowledgments

The project eHealth in Home Dialysis was part of the CleverHealth Network ecosystem and the project (eCare for Me Homedialysis) was funded by participating companies, the Hospital District of Helsinki and Uusimaa, and Business Finland. We thank our project partners Tiina Puurtinen and Pirgit Silvast-Äikäs from Helsinki University Hospital and Malin Ekholm from Helsinki University, and the patients who participated in this study.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to the ethical reasons as stated in the research permit application.

Authors' Contributions

The study was conceptualized by PV, SL, SH, NK, VR, and JV. Data were collected by five researchers (PV, SH, SL, NK, and JV). PV, SL, and SH were involved in the methodology, investigation, project administration, data curation, visualizations, and formal analysis of the study. PV finalized the experience map. JV contributed to the resources, validation, supervision, project administration, and funding acquisition of the study. VR contributed to funding acquisition, project administration, resource collection, supervision, and results' validation. All authors were involved in writing, reviewing, and editing the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Theme analysis: themes from the second phase of the analysis. CKD: chronic kidney disease.
[PDF File (Adobe PDF File), 82 KB - humanfactors_v12i1e48221_app1.pdf]

Multimedia Appendix 2
Main themes of the comprehensive affinity diagram (n=287). Some post-it notes included content for many themes.
[PDF File (Adobe PDF File), 106 KB - humanfactors_v12i1e48221_app2.pdf]

Multimedia Appendix 3
Healthy habits and hobbies, and their frequency (n=105, 18 participants).
[PDF File (Adobe PDF File), 83 KB - humanfactors_v12i1e48221_app3.pdf]

Multimedia Appendix 4
Barriers to eHealth use (n=288, 18 participants).
[PDF File (Adobe PDF File), 104 KB - humanfactors_v12i1e48221_app4.pdf]

Multimedia Appendix 5
Aspects of everyday life with chronic kidney disease, N=18 (aspects N>3 are excluded from the table)

[PDF File (Adobe PDF File), 134 KB - humanfactors_v12i1e48221_app5.pdf]

Multimedia Appendix 6
The first version of the kidney disease experience map from a user needs perspective. KD: kidney disease. CKD: chronic kidney disease.
[PNG File , 479 KB - humanfactors_v12i1e48221_app6.png]

References

1. Mendu ML, Divino-Filho JC, Vanholder R, Mitra S, Davies SJ, Jha V, International Home Dialysis Roundtable Steering Committee. Expanding utilization of home dialysis: an action agenda from the first international home dialysis roundtable. Kidney Med 2021;3(4):635-643 [FREE Full text] [doi: 10.1016/j.xkme.2021.04.004] [Medline: 34401729]
2. Liyanage T, Ninomiya T, Jha V, Neal B, Patrice HM, Okpechi I, et al. Worldwide access to treatment for end-stage kidney disease: a systematic review. Lancet 2015 May 16;385(9981):1975-1982. [doi: 10.1016/S0140-6736(14)61601-9] [Medline: 25777665]
3. Vanholder R, Annemans L, Bello AK, Bikbov B, Gallego D, Gansevoort RT, et al. Fighting the unbearable lightness of neglecting kidney health: the decade of the kidney. Clin Kidney J 2021 Jul;14(7):1719-1730 [FREE Full text] [doi: 10.1093/ckj/sfab070] [Medline: 34221379]
4. Brück K, Stel VS, Gambaro G, Hallan S, Völzke H, Ärnlöv J, European CKD Burden Consortium. CKD prevalence varies across the European general population. J Am Soc Nephrol 2016 Jul;27(7):2135-2147 [FREE Full text] [doi: 10.1681/ASN.2015050542] [Medline: 26701975]
5. GBD Chronic Kidney Disease Collaboration. Global, regional, and national burden of chronic kidney disease, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2020 Feb 29;395(10225):709-733 [FREE Full text] [doi: 10.1016/S0140-6736(20)30045-3] [Medline: 32061315]
6. Rygh E, Arild E, Johnsen E, Rumpsfeld M. Choosing to live with home dialysis-patients' experiences and potential for telemedicine support: a qualitative study. BMC Nephrol 2012 Mar 19;13:13 [FREE Full text] [doi: 10.1186/1471-2369-13-13] [Medline: 22429705]
7. Malmström RK, Roine RP, Heikkilä A, Räsänen P, Sintonen H, Muroma-Karttunen R, et al. Cost analysis and health-related quality of life of home and self-care satellite haemodialysis. Nephrol Dial Transplant 2008 Jun;23(6):1990-1996. [doi: 10.1093/ndt/gfm910] [Medline: 18223263]
8. Wu HH, Dhaygude AP, Mitra S, Tennankore KK. Home dialysis in older adults: challenges and solutions. Clin Kidney J 2023 Mar;16(3):422-431 [FREE Full text] [doi: 10.1093/ckj/sfac220] [Medline: 36865019]
9. Wang C, Ku E. eHealth in kidney care. Nat Rev Nephrol 2020 Jul;16(7):368-370 [FREE Full text] [doi: 10.1038/s41581-020-0271-z] [Medline: 32238923]
10. Hölsä S, Valkonen P, Viitanen J, Rauta V. Kohti käyttäjälähtöisiä kotidialyysin sähköisiä terveyspalveluita: Potilaskokemus ja ammattilaisten tarpeet suunnittelun lähtökohtana. FinJeHeW 2022 May 09;14(2):55. [doi: 10.23996/fjhw.114929]
11. Cornelis T, Tennankore KK, Goffin E, Rauta V, Honkanen E, Őzyilmaz A, et al. An international feasibility study of home haemodialysis in older patients. Nephrol Dial Transplant 2014 Dec;29(12):2327-2333. [doi: 10.1093/ndt/gfu260] [Medline: 25085237]
12. Muroma-Karttunen R. Almost anyone can have home therapy but can we overcome the barriers? J Ren Care 2013 Jan;39 Suppl 1:22-27. [doi: 10.1111/j.1755-6686.2013.00331.x] [Medline: 23464910]
13. Kip H, van Gemert-Pijnen L. Holistic development of eHealth technology. In: Kip H, Beerlage-de Jong N, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London, UK: Routledge; 2018:131-166.
14. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
15. Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders SM. Methods for human-centered eHealth development: narrative scoping review. J Med Internet Res 2022 Jan 27;24(1):e31858 [FREE Full text] [doi: 10.2196/31858] [Medline: 35084359]
16. ISO 9241-210:2019 Ergonomics of human-system interaction - part 210: Human-centred design for interactive systems. International Organization for Standardization (ISO). URL: https://www.iso.org/standard/77520.html [accessed 2024-04-29]
17. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. Lancet 2017 Mar 25;389(10075):1238-1252. [doi: 10.1016/S0140-6736(16)32064-5] [Medline: 27887750]
18. Development of home dialysis. CleverHealth Network. URL: https://www.cleverhealth.fi/en/development-of-home-dialysis/ [accessed 2023-01-13]
19. Gould JD, Lewis C. Designing for usability: key principles and what designers think. Commun ACM 1985 Mar;28(3):300-311 [FREE Full text] [doi: 10.1145/3166.3170]

20. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
21. van Gemert-Pijnen LJ, Kip H, Kelders SM, Sanderman R. Introducing eHealth. In: Kip H, Jong NB, van Gemert-Pijnen L, Sanderman R, Kelders SM, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. New York, NY: Routledge; 2018:3-26.
22. Brands MR, Gouw SC, Beestrum M, Cronin RM, Fijnvandraat K, Badawy SM. Patient-centered digital health records and their effects on health outcomes: systematic review. J Med Internet Res 2022 Dec 22;24(12):e43086 [FREE Full text] [doi: 10.2196/43086] [Medline: 36548034]
23. Valkonen P, Karisalmi N, Kaipio J, Kujala S. Remote interviews and visual timelines with patients: lessons learned. Stud Health Technol Inform 2021 May 27;281:845-849. [doi: 10.3233/SHTI210298] [Medline: 34042793]
24. Kushniruk AW, Borycki EM, Parush A. A case study of patient journey mapping to identify gaps in healthcare: learning from experience with cancer diagnosis and treatment. Knowl Manag E-Learn 2020;12(4):405-418 [FREE Full text] [doi: 10.34105/j.kmel.2020.12.022]
25. Ilen E, Groth C, Ahola M, Niinimäki K. Empathy in a technology-driven design process: designing for users without a voice of their own. Nordes 2019 May 10;8:1-11 [FREE Full text] [doi: 10.21606/nordes.2019.021]
26. Ector GI, Westerweel PE, Hermens RP, Braspenning KA, Heeren BC, Vinck OM, et al. The development of a web-based, patient-centered intervention for patients with chronic myeloid leukemia (CMyLife): design thinking development approach. J Med Internet Res 2020 May 15;22(5):e15895 [FREE Full text] [doi: 10.2196/15895] [Medline: 32412424]
27. Leary M, Cacchione PZ, Demiris G, Carthon JM, Bauermeister JA. An integrative review of human-centered design and design thinking for the creation of health interventions. Nurs Forum 2022 Nov;57(6):1137-1152. [doi: 10.1111/nuf.12805] [Medline: 36255150]
28. McCarthy S, O'Raghallaigh P, Woodworth S, Lim YL, Kenny LC, Adam F. An integrated patient journey mapping tool for embedding quality in healthcare service reform. J Decis Syst 2016 Jun 16;25(sup1):354-368. [doi: 10.1080/12460125.2016.1187394]
29. He Q, Du F, Simonse LW. A Patient journey map to improve the home isolation experience of persons with mild COVID-19: design research for service touchpoints of artificial intelligence in eHealth. JMIR Med Inform 2021 Apr 12;9(4):e23238 [FREE Full text] [doi: 10.2196/23238] [Medline: 33444156]
30. McCarthy S, O'Raghallaigh P, Woodworth S, Lim YY, Kenny LC, Adam F. Embedding the pillars of quality in health information technology solutions using "integrated patient journey mapping" (IPJM): case study. JMIR Hum Factors 2020 Sep 17;7(3):e17416 [FREE Full text] [doi: 10.2196/17416] [Medline: 32940610]
31. Benson M, Albanese A, Bhatia KP, Cavillon P, Cuffe L, König K, et al. Development of a patient journey map for people living with cervical dystonia. Orphanet J Rare Dis 2022 Mar 21;17(1):130 [FREE Full text] [doi: 10.1186/s13023-022-02270-4] [Medline: 35313909]
32. Kelly J, Wilden C, Herman K, Martin G, Russell C, Brown S. Botling knowledge and sharing it - using patient journey mapping to build evidence and improve aboriginal renal patent care. Ren Soc Australas J 2016;12(2):48-55 [FREE Full text] [doi: 10.3316/informit.182564296613440]
33. Joseph AL, Kushniruk AW, Borycki EM. Patient journey mapping: current practices, challenges and future opportunities in healthcare. Knowl Manag E-Learn 2020;12(4):387-404 [FREE Full text] [doi: 10.34105/j.kmel.2020.12.021]
34. Schäpers B, Krewer C, Yasuoka M, Becker D, Heller S, Kaiser M, et al. Determining design requirements for active aging: personas, experience maps, and stakeholders. Gerontechnology 2017 Oct 31;16(3):139-150 [FREE Full text] [doi: 10.4017/gt.2017.16.3.003.00]
35. Hall LK, Kunz BF, Davis EV, Dawson RI, Powers RS. The cancer experience map: an approach to including the patient voice in supportive care solutions. J Med Internet Res 2015 May 28;17(5):e132 [FREE Full text] [doi: 10.2196/jmir.3652] [Medline: 26022846]
36. Ageing Europe — looking at the lives of older people in the EU — 2020 edition. Eurostat. 2020. URL: https://ec.europa.eu/eurostat/documents/3217494/11478057/KS-02-20-655-EN-N.pdf/9b09606c-d4e8-4c33-63d2-3b20d5c19c91?t=1604055531000 [accessed 2023-01-13]
37. 11rc population according to age (5-year) and sex, 1865-2022. Statistics Finland. URL: https://pxdata.stat.fi/PxWeb/pxweb/en/StatFin/StatFin__vaerak/statfin_vaerak_pxt_11rc.px/ [accessed 2023-12-20]
38. Digital Skills Report 2023. Digital and Population Data Services Agency. URL: https://dvv.fi/documents/16079645/0/Digitaitoraportti_2023_saavutettava.pdf/4d2f74b1-0baf-a06a-f28e-954bb8df491d/Digitaitoraportti_2023_saavutettava.pdf?t=1702556379736 [accessed 2023-12-20]
39. Finnish registry for kidney diseases 2017. Munuais- ja maksaliitto. URL: https://muma.fi/wp-content/uploads/2024/02/Suomen-munuaistautirekisteri-vuosiraportti-2017.pdf [accessed 2023-12-20]
40. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]

41. Backonja U, Hall AK, Thielke S. Older adults' current and potential uses of information technologies in a changing world: a theoretical perspective. Int J Aging Hum Dev 2014 Dec;80(1):41-63 [FREE Full text] [doi: 10.1177/0091415015591109] [Medline: 26215298]
42. Latulipe C, Gatto A, Nguyen HT, Miller DP, Quandt SA, Bertoni AG, et al. Design considerations for patient portal adoption by low-income, older adults. Proc SIGCHI Conf Hum Factor Comput Syst 2015 Apr;2015:3859-3868 [FREE Full text] [doi: 10.1145/2702123.2702392] [Medline: 27077140]
43. Reiners F, Sturm J, Bouw LJ, Wouters EJ. Sociodemographic factors influencing the use of eHealth in people with chronic diseases. Int J Environ Res Public Health 2019 Feb 21;16(4):645 [FREE Full text] [doi: 10.3390/ijerph16040645] [Medline: 30795623]
44. Scheibe M, Reichelt J, Bellmann M, Kirch W. Acceptance factors of mobile apps for diabetes by patients aged 50 or older: a qualitative study. Med 2 0 2015 Mar 02;4(1):e1 [FREE Full text] [doi: 10.2196/med20.3912] [Medline: 25733033]
45. Blandford A, Furniss D, Makri S. Qualitative HCI Research: Going Behind the Scenes. Cham, Switzerland: Springer; 2016.
46. Leinonen S. Design principles of eHealth services for chronically ill patients. Master of Science Thesis, Tampere University. 2021. URL: https://trepo.tuni.fi/handle/10024/130761 [accessed 2023-01-16]
47. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, CA: Sage Publications; 2014.
48. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.
49. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
50. Kaptelinin V. Making the case for an existential perspective in HCI research on mortality and death. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. 2016 Presented at: CHI EA '16; May 7-12, 2016; San Jose, CA p. 352-364 URL: https://dl.acm.org/doi/10.1145/2851581.2892585#:~:text=The%20paper%20discusses%20some%20of,and%20evaluation%20%2D%2D%20is%20not [doi: 10.1145/2851581.2892585]
51. Kaptelinin V. Technology and the givens of existence: toward an existential inquiry framework in HCI research. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18; April 21-26, 2018; Montreal, QC p. 1-14 URL: https://dl.acm.org/doi/10.1145/3173574.3173844 [doi: 10.1145/3173574.3173844]
52. Albers R, Sadeghian S, Laschke M, Hassenzahl M. Dying, death, and the afterlife in human-computer interaction. a scoping review. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany p. 1-16 URL: https://dl.acm.org/doi/10.1145/3544548.3581199 [doi: 10.1145/3544548.3581199]
53. My Kanta pages - citizens. Kanta Services. URL: https://www.kanta.fi/en/my-kanta-pages [accessed 2023-01-16]
54. Maisa makes life easier. Client Portal Maisa.: Apotti URL: https://www.apotti.fi/en/client-portal-maisa/ [accessed 2023-01-13]
55. Kujala S, Kauppinen M, Rekola S. Bridging the gap between user needs and user requirements. In: Proceedings of the 2021 Panhellenic Conference with International Participation in Human-Computer Interaction. 2021 Presented at: PC-HCI '21; July 24-29, 2021; Virtual Event p. 45-50 URL: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=62823e8ccba3c33fa89ddb74b45c1d4c49ac878d

Abbreviations

CKD: chronic kidney disease
HCD: human-centered design
HD: home dialysis

Edited by A Kushniruk; submitted 16.04.23; peer-reviewed by N Koizumi, J Eriksen; comments to author 18.09.23; revised version received 21.12.23; accepted 15.01.25; published 18.03.25.

Please cite as:
Valkonen P, Hölsä S, Viitanen J, Leinonen S, Karisalmi N, Rauta V
Illustrating User Needs for eHealth With Experience Map: Interview Study With Chronic Kidney Disease Patients
JMIR Hum Factors 2025;12:e48221
URL: https://humanfactors.jmir.org/2025/1/e48221
doi:10.2196/48221
PMID:

Original Paper

Best Practice Guide for Reducing Barriers to Video Call–Based Telehealth: Modified Delphi Study Among Health Care Professionals

Lena Rettinger[1,2], BSc, MSc; Lea Aichinger[1], BSc, MSc; Veronika Ertelt-Bach[3], MAS; Andreas Huber[4], MSc; Susanne Maria Javorszky[5], BSc, MSc; Lukas Maul[1], BSc, MSc; Peter Putz[6], Dr rer nat; Sevan Sargis[7], BSc, MSc; Franz Werner[1], Dr. techn; Klaus Widhalm[8], MSc; Sebastian Kuhn[2], Prof Dr, MME

[1]Research Center Digital Health and Care, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria
[2]Institute of Digital Medicine, Philipps-University & University Hospital of Giessen and Marburg, Marburg, Germany
[3]Occupational Therapy, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria
[4]Orthoptics, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria
[5]Logopedics – Phoniatrics – Audiology, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria
[6]Research Center Health Sciences, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria
[7]Midwifery, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria
[8]Physiotherapy, FH Campus Wien, University of Applied Sciences Vienna, Vienna, Austria

Corresponding Author:
Lena Rettinger, BSc, MSc
Research Center Digital Health and Care
FH Campus Wien, University of Applied Sciences Vienna
Favoritenstrasse 226
Vienna, 1100
Austria
Phone: 43 6066977 ext 4382
Email: lena.rettinger@fh-campuswien.ac.at

Abstract

Background: Telehealth has grown, especially during the COVID-19 pandemic, improving access for those in remote or underserved areas. However, its implementation faces technological, practical, and interpersonal barriers.

Objective: The aim of this study was to identify and consolidate best practices for telehealth delivery, specifically for video call sessions, by synthesizing the insights of health care professionals across various disciplines.

Methods: We first identified 15 common telehealth barriers from a preceding scoping review. Subsequently, a modified Delphi method was used, involving 9 health care professionals (physiotherapists, speech and language therapists, dietitians, and midwife) with telehealth experience in qualitative interviews and 2 iterative rounds of web-based surveys to form consensus.

Results: This study addressed 15 telehealth barriers and identified 105 best practices. Among these, 20 are technology-related and 85 concern health care practices. Emphasis was placed on setting up telehealth environments, ensuring safety, building relationships and trust, using nonmanual methods, and enhancing observation and assessment skills. Best practice recommendations for dealing with patients or caregiver skepticism or lack of telehealth-specific knowledge were developed. Further, approaches for unstable networks and privacy and IT security issues were identified. Areas with fewer best practices were the lack of technology skills or technology access, unreliability of hardware and software, increased workload, and a lack of caregiver support.

Conclusions: This guide of best practices serves as an actionable resource for health care providers to navigate the complexities of telehealth. Despite a small participant sample and the potential for profession-specific biases, the findings provide a foundation for improving telehealth services and inform future research for its application and education.

(JMIR Hum Factors 2025;12:e64079) doi:10.2196/64079

KEYWORDS

telehealth; best practices; video call; Delphi study; health communication; barriers; health care professionals; qualitative interviews; web-based survey; physiotherapists; speech therapists; language therapists; dietitians; midwife

Introduction

Video call–based telehealth refers to the use of information and communication technologies to deliver health care services remotely and synchronously, bridging the gap between patients and health care providers [1]. Over the past few years, telehealth has seen an increase in adoption due to its potential to improve access to health care services, particularly for individuals living in rural or remote areas [2]. The COVID-19 pandemic has further accelerated this trend, as the need for remote health care solutions to reduce the risk of infection transmission has become more urgent [3]. As a result, telehealth has emerged as an essential part of modern health care delivery, offering new opportunities and challenges for both patients and health care providers.

Telehealth has the potential to offer numerous advantages and benefits for patients, health care providers, and the overall health care system. Benefits summarized in previous reviews include, for example, increased access and convenience, cost savings, improved family involvement, and better management [4]. Increased access refers particularly to individuals living in remote or underserved areas [2], as well as to those with mobility limitations or chronic conditions that require frequent monitoring [5]. Furthermore, telehealth enables patients to receive care from the comfort of their own homes, reducing the need for travel and waiting times associated with in-person appointments and providing a more convenient way of service provision [6]. By reducing travel costs, decreasing hospital readmissions, and minimizing the need for in-person consultations, telehealth might also have the potential to reduce health care costs for both patients and providers [7,8].

Despite its numerous advantages, telehealth also faces several challenges and barriers that can hinder its successful implementation. Recent studies and reviews have highlighted various obstacles to delivering care via real-time video consultations. These include technical issues such as unstable internet connectivity [4,9,10], lack of technology skills among both patients and providers [4,9,11], and concerns over privacy and data security [9,12]. Moreover, health care professionals have reported challenges in establishing rapport and trust with patients via video call [9,11], and there exists skepticism about the effectiveness of remote care compared to in-person sessions [9]. Addressing these barriers is especially critical when considering the rapid shift to telehealth prompted by the COVID-19 pandemic, which has underscored the potential benefits of video consultations but also exposed the need for clearer guidelines to ensure quality of care [9,13]. Despite growing evidence on these challenges, the literature lacks a comprehensive set of actionable best practices to mitigate barriers associated with video consultations. This study therefore aims to fill that gap by drawing on multiple health care disciplines and using a modified Delphi process to achieve expert consensus on best practices to maximize the potential benefits of video call–based telehealth for patients, health care providers, and the overall health care system [14].

Methods

Overview

In this study, we used a modified Delphi method [15] to identify best practices to overcome common barriers in video call–based telehealth. The reporting follows the proposed Delphi reporting guideline by Spranger et al [16].

Research Paradigm

The epistemological positioning of this study is based on a pragmatic paradigm that recognizes the value of diverse viewpoints and the importance of practical solutions to real-world problems. Pragmatism has been identified as relevant and useful for qualitative research on organizational processes [17], as well as in patient-oriented research [18]. This approach also aligns with the goals of the modified Delphi method, which aims to synthesize expert knowledge into actionable best practices. It acknowledges the dynamic and iterative nature of knowledge creation and embraces subjective experiences to address the complex challenges of telehealth implementation.

Study Design

Our modified Delphi method was characterized by 2 main adaptations to the traditional Delphi process. First, we used preliminary expert interviews to gather an initial pool of best practices. These interviews guided the development of a list of 131 practice statements and 15 barriers, which were then presented to the expert panel for rating. Second, our survey structure diverged from a classic Delphi by using (1) a 4-point Likert scale and (2) an additional response option “This practice is not relevant for my profession.” This allowed professionals from various health care disciplines to opt out of rating statements they found inapplicable. We also set specific thresholds for acceptance (≥80% agreement after round 1; ≥75% agreement after round 2) to ensure each item reached a robust consensus without necessitating a third round. These modifications provided a more streamlined approach and helped account for the diverse professional backgrounds of the participants and time and resource limitations.

Recruitment

We recruited participants via a combination of email invitations, professional social media groups, and a snowball sampling strategy. Known telehealth practitioners were first contacted directly by the research team and then invited to share the invitation within their professional networks. To ensure broad representation across allied health professions, we distributed the study invitation to professional associations and specialized digital communities. To be eligible, participants had to (1) have at least 2 years of professional experience in a non-physician health care field (eg, nursing, dietetics, occupational therapy, speech-language therapy, physiotherapy, orthoptics, or

midwifery), (2) have delivered a minimum of 10 video call–based telehealth sessions in the preceding quarter, and (3) have excellent German language proficiency. All participants were given detailed information about the study objectives, procedures, and time commitments.

Participants

A total of 9 health care professionals (4 physiotherapists (PTs), 3 speech and language therapists (SLTs), 1 dietitian, and 1 midwife) were purposively recruited based on their experience with telehealth. All participants were licensed and currently practicing health care professionals with at least 2 years of experience in their field. They each had conducted a minimum of 10 video-based telehealth sessions in the preceding quarter and practiced in an area where telehealth is applicable. Their professional experience ranged from 2 to 32 years, and encompassed diverse health care settings such as neurorehabilitation, musculoskeletal care, adults, and pediatric care, bariatric medicine, mixed practice settings, respiratory therapy, independent midwifery, and pelvic health. In total, 4 panelists were female and 5 were male.

Interviews

The expert interviews were conducted using Zoom videoconferencing software (Zoom Video Communications) in September 2022. Specifically, 4 interview sessions took place: 3 small-group focus groups and 1 individual interview. Focus group 1 included 2 PTs and lasted 1 hour 34 minutes; focus group 2 included 2 SLTs and lasted 1 hour 3 minutes; and focus group 3 included 2 PTs, 1 SLT, and 1 dietitian, lasting 1 hour and 32 minutes. The one-on-one interview was conducted with a midwife and lasted 29 minutes. We used a semistructured interview guide to help focus the discussions, which was based on commonly reported barriers with video call–based telehealth, found in the existing research for these health care fields (Multimedia Appendix 1). All interviews were conducted by the lead author (LR), who is experienced in qualitative research methods and trained in facilitating interviews and focus groups.

The barriers discussed during the expert interviews were identified through a preceding scoping review [9]: Lack of technology access, unstable internet connections, lack of privacy & IT security, lack of technology skills, unreliability of hardware and software, reluctance of patients or caregivers to use telehealth, decreased establishment of relationships and trust, decreased validity of observations and assessments, lack of hands-on methods, decreased quality and effectiveness, lack of telehealth training, increased workload, lack of support for patients from caregivers, increased safety risks for patients, and unsuitability of setting for telehealth. Experts were asked to provide insights and best practices for addressing these barriers, based on their experience.

Each interview was transcribed verbatim by the lead author (LR) with support from an automated transcription software. After automated transcription, LR reviewed the transcripts to ensure accuracy and fidelity to the original recordings. The final transcripts were analyzed using qualitative content analysis, following the framework proposed by Kuckartz and Rädiker [19]. The first author (LR) conducted the initial deductive coding by extracting passages that contained strategies for overcoming telehealth barriers. These excerpts were then assigned to the corresponding barriers. For each identified strategy, the main characteristics and essential elements were extracted to form a clear and actionable statement. To enhance the reliability of the coding, the research team engaged in multiple team rounds to discuss the initial findings. During these sessions, the team identified similarities across different strategies, combined overlapping or related strategies to reduce redundancy, and formulated precise statements that accurately captured the essence of each strategy. Through iterative discussions, codes were refined and adjusted to better represent the data, thereby enhancing the credibility of the findings.

Ethical Considerations

Prior to participation, all individuals provided informed consent, thereby ensuring ethical compliance and voluntary engagement in the study. The ethics committee confirmed the adherence to ethical principles, and an ethical approval was not required as the study involved only expert consultations without collecting or processing sensitive patient data. All participants gave informed consent, and data collection did not involve patient data. The Ethics Committee for Research Activities of FH Campus Wien, University of Applied Sciences, confirmed that no formal approval was needed for this study (EK 243/2025). The participants did not receive any financial compensation.

Delphi Rounds

Following the qualitative interviews, we conducted Delphi rounds to achieve expert consensus on best practices for video call–based telehealth. The Delphi process was carried out in 2 rounds. Round 1 was initiated at the beginning of November 2022 and remained open for 2 weeks. During this round, panelists received a web-based survey containing a demographic data section, including name and profession, and 15 sections on barriers and corresponding strategies. Round 2 was conducted at the end of November and beginning of December 2022, also open for 2 weeks. In this round, panelists were presented with a survey of 11 sections with the remaining best practices that were lacking consensus.

Barriers

Barriers were rated for their perceived magnitude (1=very small barrier, 2=rather small barrier, 3=rather big barrier, 4=very big barrier) in round 1. Frequency distributions were calculated to determine the perceived magnitude of each barrier among panelists. These results serve to contextualize the forthcoming best practice recommendations.

Analysis of Best Practices

Best practices, on the other hand, were rated for their perceived helpfulness (1=not helpful, 2=rather not helpful, 3=rather helpful, 4=very helpful). In recognition of the diverse professional backgrounds of the panelists, the survey items included an additional response option: “This practice is not relevant for my profession.” Additionally, the survey invited the panelists to contribute further comments and propose new best practices not previously covered, thereby contributing to a more comprehensive dataset.

Response percentages were calculated, and the perceived helpfulness of each practice was quantified using frequency distributions for each response category. Following the first survey round, any practice that received a "helpful" or "very helpful" rating from ≥80% of participants and elicited no comments was automatically included in the final guide. Practices with ratings between 50% and 79% or those that received substantive comments moved on to the second round for reevaluation. Regarding practices rated "helpful" or "very helpful" by <50% of participants, we weighed the risk of prematurely discarding potentially relevant items against the need to maintain focus on practices demonstrating at least a minimal level of acceptance by the panel. After careful consideration, the study team decided to exclude these practices immediately following round 1, rather than carrying them forward. However, participants were given opportunities to propose modifications or new practices at the end of each round, ensuring that any critical practice could be reintroduced or reframed if panelists believed it had merit. The research team discussed all received comments and considered if any rephrasing based on the comments should be made. In case the comments led to a rephrasing of a best practice, both the original and revised versions were presented in the second round for further evaluation. During this round, panelists were provided with the remaining best practices along with the percentage of previous agreement and anonymized comments for each best practice in December 2022. Participants were offered 2 response options: "This practice should be included in the best practice guide" or "This practice should not be included in the best practice guide." When 2 variations of a practice were presented, panelists could select their preferred version or "none of both." Practices that carried over from round 1 required at least 75% agreement among participants to be included in the final best practice guide [20]. This slightly lower threshold accounted for the fact that these practices already showed moderate support or warranted further discussion due to comments. In case 2 versions of a best practice existed, the version receiving the higher preference was selected.

Results

Barriers and Best Practices

The qualitative content analysis on the expert interviews yielded a total of 422 coded text segments. From this process, an initial set of 131 practices was developed. The association of the practices with the discussed barriers can be found in Table 1, and the perceived importance of the 15 presented barriers is shown in Figure 1.

In the first Delphi round, 108 practices exceeded the 80% approval threshold for perceived helpfulness; 83 of these practices had no comments and were therefore included directly. In total, 48 practices met the criteria to proceed to the second round: 23 practices rated as "very helpful" or "helpful" by 50%-79% of the experts; 18 without changes but with comments from the first round; and 7 practices with alternative revision based on comments. One entirely new strategy was introduced (Figure 2). Upon completion of the second round, the final best practice guide emerged with 105 practices, covering practices addressing 5 technology barriers and 10 practical barriers (Table 2 and Textbox 1).

In total, we identified 105 best practices addressing 15 telehealth barriers, with 20 aimed at technology challenges and 85 focused on practice issues.

Table 1. Demographic characteristics of the panelists.

Panelist number	Gender	Profession	Years of professional experience	Professional experience
1	Male	Physiotherapist	15	Orthopedics, traumatology, neurolrehabilitation
2	Male	Physiotherapist	3	Orthopedics, musculoskeletal, training
3	Female	Speech and language therapist	20	Swallowing disorders, functional voice disorders
4	Female	Speech and language therapist	2	Childhood speech development disorders
5	Male	Physiotherapist	32	Orthopedics, respiratory therapy
6	Female	Dietitian	4	Bariatrics
7	Male	Speech therapist	2.5	General speech therapy, respiratory therapy
8	Female	Physiotherapist	20	Pelvic floor therapy
9	Female	Midwife	28	Birth preparation, support, and aftercare

Figure 1. Panelists' (n=9) perceived magnitude of the 15 barriers presented (bars represent absolute frequencies).

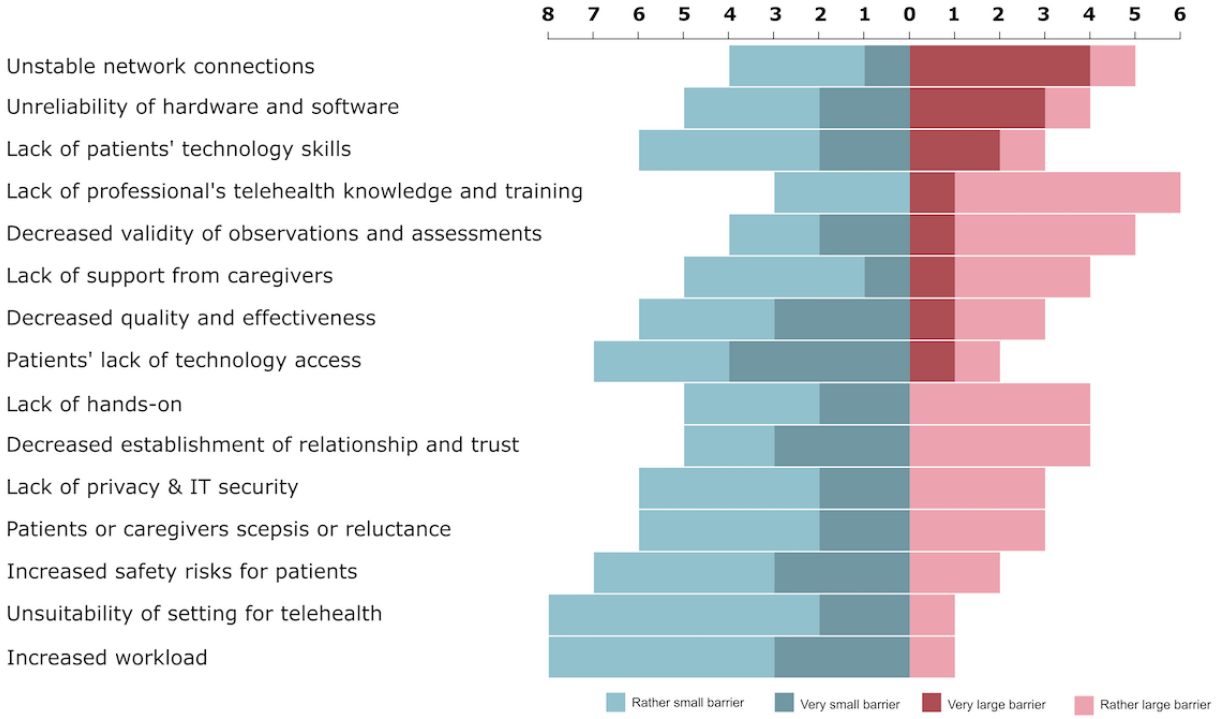


Figure 2. Flowchart of the Delphi process.

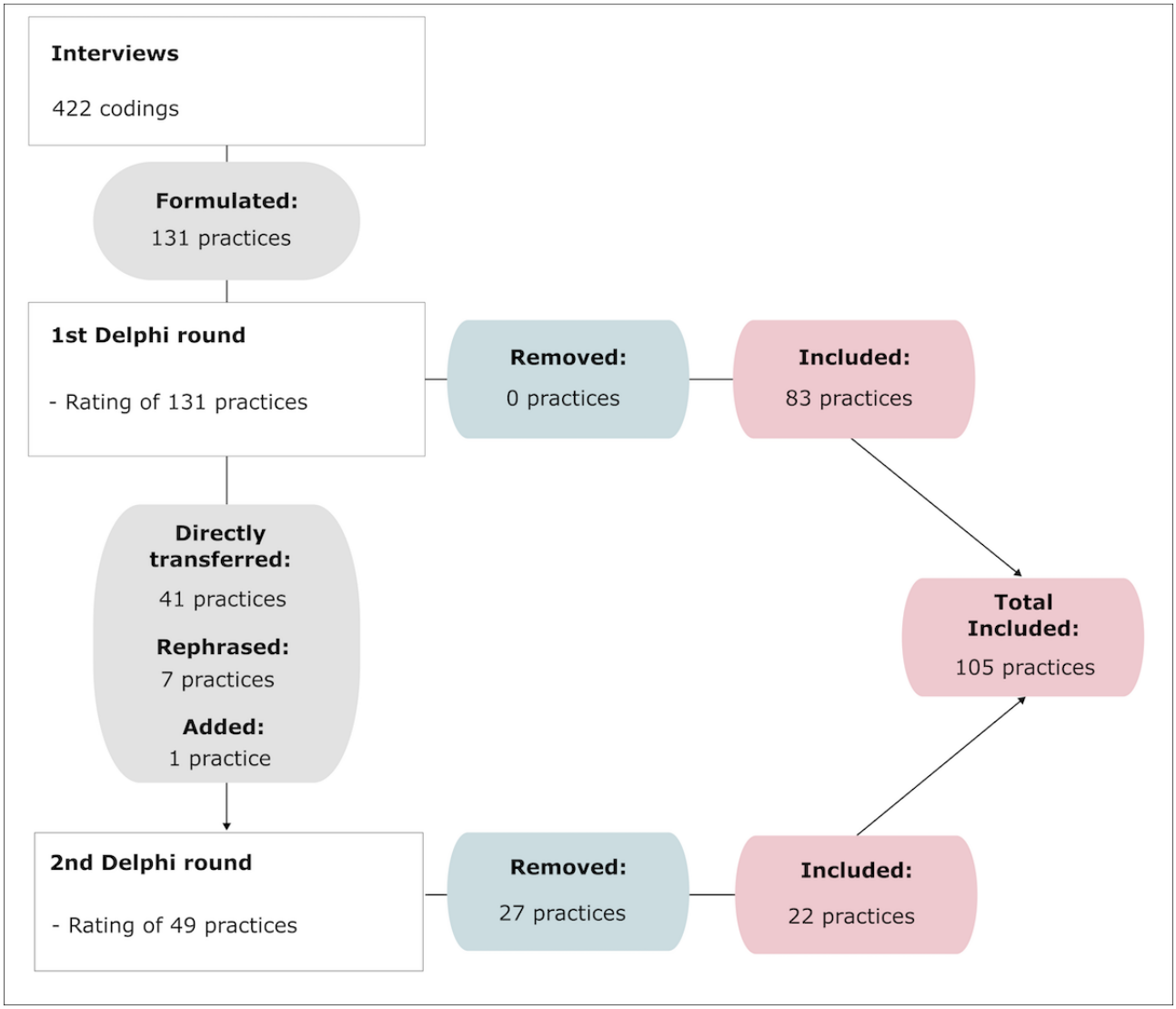


Table 2. Number of best practices for each presented barrier.

Barrier	Number of practices after expert interviews	Number of practices, included after round 1	Number of practices included after round 2	Final number of practices
Technology barriers				
Unstable network connections	9	3	3	6
Unreliability of hardware and software	4	1	1	2
Lack of technology skills	6	2	2	4
Lack of privacy and IT security	7	2	3	5
Lack of technology access	3	0	3	3
Health care practice barriers				
Lack of professional's telehealth knowledge and training	8	6	0	6
Decreased validity of observations and assessments	13	8	2	10
Lack of support from caregivers	3	1	0	1
Lack of hands-on	11	7	3	10
Reluctance of patients or caregivers to use telehealth	8	5	2	7
Decreased establishment of relationship and trust	11	9	1	10
Decreased quality and effectiveness	10	9	1	10
Increased safety risks for patients	13	10	1	11
Unsuitability of setting for telehealth	21	18	0	18
Increased workload	4	2	0	2
Sum total	131	83	22	105

Textbox 1. Best practice guide for barriers to telehealth via video call.

Best practices to address technology barriers

A. Unstable network connections

1. Every participant should choose a location with a stable network connection for video call sessions. This should be communicated to the clients in advance.
2. Other devices, persons, or activities should not use the same network connection during the video call session.
3. In case of an unstable network connection, the device may be exchanged.
4. In case of an unstable network connection, switching off the camera for a short time could help.
5. In case of network disconnection, the client should be informed about further steps by telephone (full agreement from all participants).
6. In case of audio transmission problems, an additional phone connection can be used.

B. Unreliability of hardware and software

1. The video call software used should be installed and tested in a timely manner by all participants before the start of the session.
2. If clients experience technical issues, possible solutions can be discussed and solved directly during the video call session or via telephone (full agreement from all participants).

C. Lack of technology skills

1. Health care professionals should gain digital literacy skills in using different devices and video call software, as this can help in solving common technical problems encountered by patients (full agreement from all participants).
2. For technically non-proficient individuals, user-friendly video call software should be selected. Getting started via a direct link without prior installation should be favored (full agreement from all participants).
3. Telephone support can help inexperienced individuals get started with video call software (full agreement from all participants).
4. As required, the operation of the video call software can be supported at the beginning or ongoing by relatives or other persons on site.

D. Lack of privacy & IT security

1. Software used should ensure encrypted data transmission (full agreement from all participants).
2. One's own decisions regarding the handling of client data should be made and communicated in accordance with current data protection laws.
3. Written consent from clients should be obtained in advance.
4. When using video call software, access should only be possible for invited individuals (full agreement from all participants).
5. If software has not been recommended by a certification body for use in health care, it should be reviewed concerning its data processing.

E. Lack of technology access

1. Providing a checklist of minimum technical requirements for video call sessions can optimize conditions in advance.
2. Availability of equipment should be confirmed prior to the beginning of the session.
3. For sessions that focus on visual demonstration or perception, it is recommended to use a device with a large, high image quality screen.

F. Lack of telehealth knowledge and training

1. Watching video call software tutorials can provide information on available functions and give confidence in handling.
2. Learning by doing can help to gain confidence in conducting video call sessions.
3. A test run with individuals from the private environment can provide confidence in using the technology (full agreement from all participants).
4. Materials for use in video call sessions can be researched on the internet (full agreement from all participants).
5. Reading current articles on the topic of telehealth can increase knowledge (full agreement from all participants).
6. Exchanging experiences with colleagues about the use of video units can expand one's own repertoire (full agreement from all participants).

G. Decreased validity of observations and assessments

1. Important assessments should be conducted in a supplementary face-to-face unit if specific information cannot be gathered via a video call session.
2. Clients should be informed that their clothing of choice should allow detailed observations of the body (full agreement from all participants).
3. Cameras should be positioned in a way that enables a stable image (full agreement from all participants).

4. Detailed and guided instructions on how to position the body or involved body parts in front of the camera should be given to ensure high-quality observations (full agreement from all participants).
5. The background should provide visual reference points for guiding and observing movements.
6. The observation quality of details can be improved by moving the corresponding body section closer to the camera (full agreement from all participants).
7. The observation quality of details can be improved with targeted lighting (full agreement from all participants).
8. Assessments can be adapted with the help of using objects in the client´s room during video call sessions (full agreement from all participants).
9. Increased attention should be given to small nonverbal signals of the clients.
10. Involving self-observation and self-perception of clients can supplement limited on-screen observation possibilities (full agreement from all participants).

H. Lack of support from caregivers

1. Ensuring that a support person is within reach can help to run a video call session smoothly (full agreement from all participants).

I. Lack of hands-on

1. Education about functional deficits and how to deal with them in everyday life should be emphasized in video call sessions.
2. Training self-awareness and empowering clients in their self-efficacy can be part of video call sessions to achieve lasting effects (full agreement from all participants).
3. Proven effectiveness of therapy interventions aiming at lifestyle modifications should be explained to clients (full agreement from all participants).
4. Exercises that can be performed by the clients independently should be selected (full agreement from all participants).
5. Exercises should be guided verbally in such a manner that they can be understood even if visual guidance is limited (full agreement from all participants).
6. In advance, material needed for the video call session should be communicated to the clients or relatives, and it should be prepared in a timely manner (full agreement from all participants).
7. Selected self-treatments can be taught to clients during the video call session (full agreement from all participants).
8. Clients should be given specific advice for performing self-control of guided exercises (full agreement from all participants).
9. Clients' self-awareness can be promoted by feeling movements or structures with their own hands, wherein possible misinterpretations should be considered.
10. Material required for a video call session should be communicated in advance with clear information, for example, on size, quantity, and weight.

J. Patients or caregivers scepsis or reluctance

1. Clients can be offered a combination of presence and video call sessions (full agreement from all participants).
2. It can be explained to skeptical clients that some therapy content is suitable for video call sessions, others not (full agreement from all participants).
3. Demonstrating video call software and discussing a possible video call session can reduce insecurities and help with decision-making.
4. A single video call session for familiarization can be offered to skeptical clients, followed by a discussion about the experience and further procedure.
5. After a few face-to-face sessions, video call sessions can be suggested again.
6. Involving relatives or the social environment can be facilitated with video call sessions.
7. The reduction of travel and increase in time flexibility can be used as an argument for using video call sessions.

K. Decreased establishment of relationship and trust

1. Authenticity and professionalism can be conveyed through conscious design choices and insight into the premises in the background (full agreement from all participants).
2. Optimally adjusted lighting of one's face can create a friendly atmosphere and convey mimic information well (full agreement from all participants).
3. A well-adjusted camera angle and optimal positioning in front of the screen support communication on an equal footing (full agreement from all participants).
4. Facial expressions and gestures during a video call session should be used consciously (full agreement from all participants).
5. Since real eye contact is not possible via the screen, this can be imitated by consciously looking into the camera.
6. Maintaining one's own appearance can promote authenticity in front of the camera (full agreement from all participants).
7. Conscious choice of clothing supports the professional appearance.

8. Seeing one's own camera image can be used as feedback about one’s own appearance (full agreement from all participants).
9. Framing a video call session with introductory words and a deliberate end can provide orientation (full agreement from all participants).
10. Spontaneous playful elements and humor can loosen up video call sessions and promote the relationship (full agreement from all participants).

L. Decreased quality and effectiveness

1. Clients who can benefit from video call sessions should be selected purposefully (full agreement from all participants).
2. The limitations of video call sessions for certain medical conditions should be clearly communicated to clients (full agreement from all participants).
3. The use of video call sessions should be evaluated regularly and adjusted if necessary.
4. Strategies to ensure whether elements learned in the session can be implemented independently should be used.
5. Video call sessions can be used to maintain a regular appointment frequency (full agreement from all participants).
6. By collecting relevant health parameters at the beginning and the end of the session, the effectiveness of treatment interventions can be evaluated.
7. At the beginning of a video call session, expectations should be gathered. At the end they should be compared with the session's content and success.
8. The role of relatives in the treatment process can be strengthened through targeted counseling during video call sessions.
9. By setting a clear focus and sequence for the video call session and individualizing the duration, one’s ability to focus can be considered (full agreement from all participants).
10. The video call session should take place during a time of the day when clients have a good ability to focus (full agreement from all participants).

M. Increased safety risks for patients

1. Clear exclusion and termination criteria for video call sessions for clients with safety-critical characteristics should be established and adhered to (full agreement from all participants).
2. Competent relatives can increase clients’ safety.
3. For video call sessions with children, a legal guardian should be present at the child's side (full agreement from all participants).
4. Clients or family members should be educated about making the environment as safe as possible to reduce potential hazards such as tripping, falling, or slipping (full agreement from all participants).
5. Client’s pets that may interfere with free movement in the room should be kept out of the client's room during video call sessions.
6. Exercises instructed via video call sessions should be chosen according to clients' abilities (full agreement from all participants).
7. Video call sessions should not include unfamiliar exercises that may jeopardize client safety.
8. Exercise difficulty should be increased only step by step, adapted to the client (full agreement from all participants).
9. In case of uncertainties regarding perceived health parameters, referral to specialists should be made (full agreement from all participants).
10. In psychologically tense situations, one’s own voice can serve as an instrument for calming down (full agreement from all participants).
11. The daytime of the video call session, and therefore a potentially limited suitability of certain measures, should be taken into consideration.

N. Unsuitability of setting for telehealth

1. A quiet room without distractions should be chosen for video call sessions (full agreement from all participants).
2. Background noise from people, pets, or household equipment should be reduced (full agreement from all participants).
3. For video call sessions, a room with good sound quality and reduced hall should be chosen (full agreement from all participants).
4. The speaker used should be tuned sufficiently loud (full agreement from all participants).
5. Balanced lighting conditions allowing good contrasts should be selected (full agreement from all participants).
6. Devices should be positioned stable to not produce blurry images (full agreement from all participants).
7. If a tablet or smartphone is used, it should be positioned at a 90-degree angle to the surface with a suitable mount (full agreement from all participants).
8. Clients should be guided in the optimal camera setup and positioning in front of it (full agreement from all participants).
9. The camera section should be chosen so that only things that one is willing to share are visible in the background (full agreement from all participants).
10. Other people in the household or work premises should be informed that a video call session is taking place to ensure privacy (full agreement from all participants).
11.

It should be discussed with the clients in which area of their premises the video session will take place in order to avoid the concern of giving insight into things they want to keep private (full agreement from all participants).

12. Windows and doors should be closed.
13. High-quality microphones and headphones can improve acoustic transmission.
14. Advice should be given on how to use built-in or external microphones to ensure good transmission and avoid noise interference.
15. Doors in the camera's field of view should be avoided to prevent unwanted privacy intrusion.
16. A dedicated therapy location—for example, at a table—can promote attention and concentration.
17. Marking one or more places where the camera is well positioned facilitates a sustainably working set-up.
18. If the video call session involves movement across the room, a flexible camera setup for rapid changes should be facilitated.

O. Increased workload

1. Using quick-to-use treatment or exercise methods that require little pre-planning and explanation can reduce preparation time.
2. By combining preparation time for several video units, one's own working time can be used efficiently.

Best Practices to Address Technology Barriers

In total, 20 best practices were selected by the participants to address the barriers of unstable network connections, unreliability of hardware and software, lack of technology skills, lack of privacy and IT security, and lack of technology access (A-E in Textbox 1). For optimal video call sessions, each participant should ensure a stable network connection, free from other devices and activities that might compete for bandwidth. When issues such as an unstable connection arise, switching devices, temporarily turning off the camera, or making a phone call can provide continuity and effective communication. Participants should install and test the video call software well before the session, with technical issues being addressed directly during the call or via telephone if necessary. Health care professionals should develop digital literacy for various devices and software to assist patients with technical issues. For less tech-savvy individuals, user-friendly software with direct links, telephone support, as well as on-site help from relatives, can facilitate the use of video call applications. To ensure data privacy and security, video call software must encrypt data transmission and comply with data protection laws. Informed consent should be obtained from users, and access should be restricted to invited participants. If software has not been recommended by a certification body for use in health care, it should be reviewed concerning its data processing. To optimize video call sessions, a checklist of technical requirements should be provided, and equipment availability should be confirmed beforehand. For visually intensive sessions, a high-quality large screen is recommended.

Best Practices to Address Health Care Practice Barriers

Overall, 85 best practices addressed practice issues such as lack of telehealth knowledge and training, decreased validity of observations and assessments, lack of support from caregivers, lack of hands-on, patients or caregivers scepsis or reluctance, decreased establishment of relationship and trust, decreased quality and effectiveness, increased safety risks for patients, unsuitability of setting for telehealth, and increased workload (F-O in Textbox 1).

Reviewing current telehealth publications and internet resources to search for material to use in telehealth sessions, as well as watching software tutorials, can enhance proficiency and confidence in conducting video call sessions, complemented by practical experience and sharing knowledge with colleagues.

For comprehensive assessments in telehealth, certain procedures may need to be supplemented with in-person visits. Clients should be guided on how to prepare for video sessions with appropriate clothing, camera positioning, lighting, and background setup to enhance the quality of observations. Additionally, using objects for adapted assessments, paying attention to subtle nonverbal cues, and involving clients in self-observation can maximize the efficacy of remote evaluations. Having a support person readily available can facilitate a smooth video call session.

In video call sessions, educating clients on functional deficits and self-management in daily life is crucial, with a focus on training self-awareness and self-efficacy for sustainable outcomes. For effective video call sessions, exercises that clients can perform independently should be chosen, and clear verbal guidance for exercises with limited visual cues should be given. This can be added with the training of self-treatments, advice for self-control during exercises, and the promotion of clients' self-awareness through tactile experiences while addressing potential misinterpretations. Clear communication and preparation of materials is necessary.

To address skepticism, clients can be offered a blend of in-person and video call sessions, emphasizing the suitability of certain therapy contents for remote delivery, with demonstrations and trial sessions helping to alleviate skepticism and showcasing the convenience and flexibility that video calls offer. To convey professionalism and authenticity in video calls, one should make deliberate design and setting choices, ensure optimal lighting for a friendly atmosphere, and use a camera angle that supports equitable communication. Conscious use of facial expressions, gestures, and clothing, along with looking into the camera to simulate eye contact, enhances authenticity. Additionally, framing the session with a clear beginning and end, along with the use of humor, can improve the interaction.

Purposeful selection of clients for video call sessions is essential, with transparent communication regarding their limitations for certain conditions. Regular evaluations should be conducted to assess effectiveness and determine if expectations are being met. Strategies to ensure whether elements learned in the session can be implemented independently should be used. Sessions should be structured in a way that they have a clear focus and sequence, an appropriate duration, and are scheduled for when the client's focus can optimize the session. Incorporating relatives into the treatment process via video call sessions can provide targeted counseling. Establishing clear criteria for excluding or terminating video call sessions is vital for clients with safety-critical concerns. Competent relatives can enhance safety. Guardian presence for children, safety education, and gradual exercise progression can ensure the clients' well-being during remote sessions.

For effective video call sessions, it is essential to choose a quiet room with as little background noise as possible, ensuring good sound quality and balanced lighting. Proper camera setup, privacy considerations, and guidance on microphone use enhance the session quality. Additionally, creating a dedicated therapy location and facilitating flexible camera setups can improve the overall experience and engagement during sessions. Efficient use of time in video sessions can be achieved by opting for quick and easy treatment methods and combining preparations for multiple video units.

Discussion

Principal Findings

The findings from this study, which encompassed expert interviews and a modified Delphi process, provide a best practice guideline for telehealth in health care. The key results underscore the importance of addressing the challenges of technology accessibility, enhancing observations and assessments, creating a professional setup, building relationships and trust, and ensuring privacy and IT security. Although the majority of the panelists were PTs and SLTs, their combined professional experiences—spanning areas such as neurorehabilitation and pediatric care—have facilitated a comprehensive exploration of these practices across various health care settings, thereby complementing the aims and findings of previous studies and guidelines [21-25].

The best practice guideline offers a starting point towards a unified guide to improve telehealth across various health care settings. Prior publications have established telehealth frameworks in disciplines such as occupational therapy, physiotherapy, and speech and language therapy [21-24], which partially align with our findings. A high level of agreement is seen in these existing frameworks, particularly regarding fundamental principles of telehealth delivery. However, there remain clear gaps in the current literature on how to manage these barriers through concise best practice statements [24]. Grogan-Johnson [25] discusses the challenges of maintaining engagement in virtual settings but provides limited guidance on how to address this. Our best practice guide suggests trust-building strategies, such as non-verbal communication enhancements, environmental adjustments, and interactive session structuring. Davies et al [13] outline the importance of technological proficiency for PTs but do not provide detailed practices for managing unstable networks or unreliable hardware, where our study adds explicit recommendations for overcoming technological challenges, including strategies for low-bandwidth settings, selecting robust hardware, and providing patient education on basic troubleshooting. Other studies limit their focus to specific populations or pathologies [22,25]. Cottrell and Russell [22], for instance, provide an in-depth analysis of pre-implementation considerations and practical steps specifically within the domain of musculoskeletal physiotherapy. Their work examines technological and operational challenges, offering targeted strategies to enhance the successful deployment of telehealth services. Other guidelines have delineated key technical principles health care professionals are expected to follow to ensure a smooth and secure telehealth experience [26,27], focusing on adherence to laws and regulations and the implementation of privacy and security measures. Our study extends these principles by proposing practical, applicable strategies and emphasizes dual-focused education, targeting both health care providers and patients. It includes specific methods to address skepticism and enhance digital literacy, particularly for populations unfamiliar with telehealth.

Notably, our research revealed a strong consensus on the effectiveness of specific strategies across health care settings, such as providing detailed technical guidelines, enhancing visual and audio components of telehealth sessions, and incorporating elements that foster a sense of connection and trust between health care providers and clients. These findings offer new insights into the practical implementation and adaptation of telehealth services in diverse health care contexts.

In examining the distribution of best practices identified to address various barriers, we found notable differences. "Unsuitability of setting for telehealth" yielded the largest set of best practices despite being rated as a relatively small barrier. This contrast likely arose because many aspects of the setting (eg, lighting, device positioning, noise reduction) are easily modified by health care professionals. These direct, practitioner-driven solutions help frame the barrier as "small," in that it can be quickly addressed through practical steps. In other words, the perception of a barrier's magnitude may reflect how empowered providers feel to remedy it: settings are simpler to adjust, whereas more systemic issues (eg, workload, limited technology access) require external interventions and may thus seem more significant.

Additionally, a significant collection of best practices has been identified to address practical implementation barriers such as "increased safety risks for patients," "decreased validity of observations and assessments," "lack of hands-on," "decreased establishment of relationship and trust," and "decreased quality and effectiveness." These best practices provide pragmatic strategies to navigate the range of challenges presented by telehealth, ensuring a higher standard of patient care and treatment integrity. On the other side, there were barriers that received fewer best practices to address them, such as "lack of support from caregivers" and "increased workload," both at the lower end, with 1 and 2 identified best practices. These issues

likely require organizational or systemic changes that are beyond the immediate control of the individual practitioner. Similarly, "lack of technology access" and "unreliability of hardware and software" are countered by only 3 best practices each, possibly indicating a gap between the technological requirements of telehealth and the ability of health professionals to influence these parameters.

Reflections on the Delphi Method

Our modified Delphi approach relied on 2 rounds, underpinned by preliminary interviews that generated a well-defined set of barriers and best practices. This groundwork contributed to a high level of initial consensus—83 of 108 items achieved ≥80% agreement in round 1. While we acknowledge that time constraints and difficulty finding sufficient participants limited the scope of our panel (n=9), each participant brought substantial experience: all met strict inclusion criteria (eg, minimum 2 years' professional practice, at least 10 video-based telehealth sessions in the previous quarter), ensuring specialized and recent expertise.

We employed a 2-threshold system: items achieving ≥80% agreement in round 1 were immediately included, while items carried forward required ≥75% agreement in round 2. This tiered approach balanced efficiency and adaptability by refining practices with moderate support without burdening participants with repeated rounds. Items garnering <50% agreement were excluded after round 1, given their limited traction; however, participants could introduce revisions or propose new items if they felt critical information was overlooked. Although this may appear stringent, it enabled us to produce a focused, consensus-driven guide for reducing barriers to video call–based telehealth within the time and resource constraints of our project.

While our study does not claim to capture every possible facet of telehealth across all health care contexts, the methodology employed aligned with recommended standards.

Limitations

This study still has several limitations that need to be acknowledged. First, the sample size was small, consisting of only 9 eligible experts in specific health professions. Our participant sample was relatively small, yet it was purposefully recruited to encompass diverse professional backgrounds and extensive telehealth experience. This approach generated a robust consensus on best practices for video call–based telehealth. Future studies with larger or more varied panels could further enhance and validate the findings, but our current methods provide a strong foundation for ongoing research and practical implementation. Second, the heterogeneity of the participants' professions may be both an advantage and a disadvantage. On the one hand, it allowed for a broader range of perspectives and strategies to be considered than if we would have focused on only one profession. On the other hand, it could also have resulted in the exclusion of profession-specific strategies that may be essential for a successful telehealth implementation, which could be studied in the future. Each profession uses telehealth differently based on their specific clinical needs and methods. For instance, PTs may focus more on movement assessments and exercise guidance, whereas SLTs might emphasize communication techniques and auditory assessments. This variation means that certain best practices identified may be more applicable to some professions than others, potentially limiting the universal applicability of the recommendations. The small number of distinct professions included in the study may have resulted in the exclusion of specialized best practices pertinent to each field. Essential strategies unique to specific clinical scenarios or patient interactions within each profession might not have been fully captured or emphasized, thereby affecting the comprehensiveness of the best practice guide for individual disciplines.

Moreover, the findings reflect the current state and must be interpreted as such. They likely are subject to change due to the continuous evolution of professional expertise in health care and ongoing technological advancements that are likely to shape the future landscape of telehealth. Furthermore, this study relied on self-reported data from the experts. While this method is commonly used in qualitative research, it is important to acknowledge that the data collected may be subject to bias or limitations in recall. Another important consideration is the length of the final best practice list. Although it represents a comprehensive set of potential solutions, its sheer number may pose challenges for day-to-day clinical application.

Implication of Findings

The study's findings offer valuable insights for various stakeholders within the telehealth ecosystem, promoting enhanced effectiveness, accessibility, and sustainability of video call–based telehealth services. The best practices serve as a practical guide for clinicians to optimize telehealth delivery. Implementing these strategies can improve care quality, strengthen patient-provider relationships, and increase patient satisfaction while reducing operational burdens through streamlined session management and technical troubleshooting. Using enhanced privacy measures and user-friendly technologies ensures a secure and confident telehealth experience for patients, particularly those in remote or underserved areas. Insights from this study can also inform the development of regulations and standards that promote equitable telehealth access. By addressing practical barriers and supporting measures such as technology funding and training programs, policy makers can facilitate the broader adoption and integration of telehealth into standard health care practices. Finally, incorporating these best practices into health care curricula ensures that future practitioners are equipped with the necessary digital literacy and remote communication skills. Training programs can emphasize technical competencies and patient engagement strategies specific to telehealth, preparing health care workers for modern care delivery.

Further Research

While this Delphi study provided insights into potential solutions, it would be beneficial to evaluate the impact of these strategies on the successful implementation of telehealth in different real-world settings by exploratory and confirmatory studies. For example, while this study focused on barriers and solutions in the selected health care professions, other health care specialties and settings may face different challenges and

would come up with different best practices when implementing telehealth. Evaluating these barriers and solutions in different contexts would provide a more comprehensive understanding of telehealth implementation. Moving forward, further research could investigate strategies to streamline or prioritize these recommendations, identifying which are most beneficial for specific professional groups or patient populations. Further, it would be interesting to create educational programs to see how teaching these strategies to students would increase their telehealth proficiency.

Conclusions

This publication presents a modified Delphi study that compiles best practices for overcoming barriers in video call–based telehealth. It includes insights from various health care professionals and focuses on enhancing telehealth's quality and effectiveness. The study emphasizes technology accessibility, professional setups, and patient-centered care. Limitations include a small sample size and self-reported data, suggesting future research should assess these strategies' real-world effectiveness, explore barriers in different health care specialties, and develop educational programs to increase telehealth proficiency.

Acknowledgments

The authors wish to thank the participants of the Delphi study. This work was supported by the City of Vienna, Magistratsabteilung 23, Austria (grant MA23-338474-2021-2).

Data Availability

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

Authors' Contributions

LR contributed to project administration; LR and LA were involved in conceptualization, supported by all authors; LR and PP contributed to methodology; data curation was done by LR; formal analysis and investigation were done by LR and LM; LR involved in writing—original draft preparation; all authors contributed to writing—review and editing; funding acquisition was done by LR and FW; supervision was done by SK.

Conflicts of Interest

SK is the founder and shareholder of MED.digtal. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Interview Guideline_Translated version_English.
[PDF File (Adobe PDF File), 63 KB - humanfactors_v12i1e64079_app1.pdf]

References

1. Mechanic OJ, Persaud Y, Kimball AB. Telehealth Systems. StatPearls Treasure Island, FL: StatPearls Publishing; 2022.
2. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. PLoS One 2019;14(8):e0221848 [FREE Full text] [doi: 10.1371/journal.pone.0221848] [Medline: 31469865]
3. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
4. Almathami HKY, Win KT, Vlahu-Gjorgievska E. Barriers and facilitators that influence telemedicine-based, real-time, online consultation at patients' homes: systematic literature review. J Med Internet Res 2020;22(2):e16407 [FREE Full text] [doi: 10.2196/16407] [Medline: 32130131]
5. Gellis ZD, Kenaley B, McGinty J, Bardelli E, Davitt J, Ten Have T. Outcomes of a telehealth intervention for homebound older adults with heart or chronic respiratory failure: a randomized controlled trial. Gerontologist 2012;52(4):541-552. [doi: 10.1093/geront/gnr134] [Medline: 22241810]
6. Kruse CS, Krowski N, Rodriguez B, Tran L, Vela J, Brooks M. Telehealth and patient satisfaction: a systematic review and narrative analysis. BMJ Open 2017;7(8):e016242 [FREE Full text] [doi: 10.1136/bmjopen-2017-016242] [Medline: 28775188]
7. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]
8. Snoswell CL, Smith AC, Page M, Scuffham P, Caffery LJ. Quantifying the societal benefits from telehealth: productivity and reduced travel. Value Health Reg Issues 2022;28:61-66 [FREE Full text] [doi: 10.1016/j.vhri.2021.07.007] [Medline: 34800833]

9. Rettinger L, Kuhn S. Barriers to video call-based telehealth in allied health professions and nursing: scoping review and mapping process. J Med Internet Res 2023;25:e46715 [FREE Full text] [doi: 10.2196/46715] [Medline: 37526957]
10. Van Eerdenbrugh S, Schraeyen K, Leysen H, Mostaert C, D'haenens W, Vandenborre D. Delivery of speech-language therapy and audiology services across the world at the start of the COVID-19 pandemic: a survey. Perspect ASHA Spec Interest Groups 2022;7(2):635-646. [doi: 10.1044/2021_persp-21-00134]
11. Rosenstrøm S, Groth S, Risom S, Hove J, Brødsgaard A. Nurses' experiences with virtual consultations and home-monitoring in patients with cardiac disease: a systematic review and qualitative meta-synthesis of results. Nurs Health Sci 2024;26(4):e13180. [doi: 10.1111/nhs.13180] [Medline: 39437992]
12. Greenhalgh T, Shaw S, Wherton J, Vijayaraghavan S, Morris J, Bhattacharya S, et al. Real-world implementation of video outpatient consultations at macro, meso, and micro levels: mixed-method study. J Med Internet Res 2018;20(4):e150 [FREE Full text] [doi: 10.2196/jmir.9897] [Medline: 29625956]
13. Davies L, Hinman RS, Russell T, Lawford B, Bennell K, International Videoconferencing Steering Group. An international core capability framework for physiotherapists to deliver quality care via videoconferencing: a Delphi study. J Physiother 2021;67(4):291-297 [FREE Full text] [doi: 10.1016/j.jphys.2021.09.001] [Medline: 34521586]
14. Jennett PA, Andruchuk K. Telehealth: 'real life' implementation issues. Comput Methods Programs Biomed 2001;64(3):169-174. [doi: 10.1016/s0169-2607(00)00136-x] [Medline: 11226614]
15. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021;11(4):116-129 [FREE Full text] [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]
16. Spranger J, Homberg A, Sonnberger M, Niederberger M. Reporting guidelines for Delphi techniques in health sciences: a methodological review. Z Evid Fortbild Qual Gesundhwes 2022;172:1-11 [FREE Full text] [doi: 10.1016/j.zefq.2022.04.025] [Medline: 35718726]
17. Kelly LM, Cordeiro M. Three principles of pragmatism for research on organizational processes. Methodol Innov 2020;13(2):205979912093724. [doi: 10.1177/2059799120937242]
18. Allemang B, Sitter K, Dimitropoulos G. Pragmatism as a paradigm for patient-oriented research. Health Expect 2022;25(1):38-47 [FREE Full text] [doi: 10.1111/hex.13384] [Medline: 34748689]
19. Kuckartz U, Rädiker S. Qualitative Content Analysis: Methods, Practice and Software. Thousand Oaks: SAGE Publications; 2023.
20. Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. J Clin Epidemiol 2014;67(4):401-409. [doi: 10.1016/j.jclinepi.2013.12.002] [Medline: 24581294]
21. Almubark BM, Majrashi N, Alghusun N, Alhammad M, Alhthifi F, Alyahya RS. Telehealth clinical practice guide for occupational therapy, physical therapy, and speech and language pathology: a Saudi and middle eastern guide. Telemed J E Health 2022;28(5):636-642. [doi: 10.1089/tmj.2021.0021] [Medline: 34529497]
22. Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy. Musculoskelet Sci Pract 2020;48:102193 [FREE Full text] [doi: 10.1016/j.msksp.2020.102193] [Medline: 32560876]
23. World Federation Of Occupational Therapists. World federation of occupational therapists' position statement on telehealth. Int J Telerehabil 2014;6(1):37-39 [FREE Full text] [doi: 10.5195/IJT.2014.6153] [Medline: 25945221]
24. Cason J, Hartmann K, Jacobs K, Richmond T. Telehealth in occupational therapy. Am J Occup Ther 2018;72(Supplement_2):7212410059p1-7212410059p18. [doi: 10.5014/ajot.2018.72S219] [Medline: 30674404]
25. Grogan-Johnson S. The five W's meet the three R's: the who, what, when, where, and why of telepractice service delivery for school-based speech-language therapy services. Semin Speech Lang 2021;42(2):162-176 [FREE Full text] [doi: 10.1055/s-0041-1723842] [Medline: 33725733]
26. Anil K, Freeman JA, Buckingham S, Demain S, Gunn H, Jones RB, et al. Scope, context and quality of telerehabilitation guidelines for physical disabilities: a scoping review. BMJ Open 2021;11(8):e049603 [FREE Full text] [doi: 10.1136/bmjopen-2021-049603] [Medline: 34385253]
27. Leone E, Eddison N, Healy A, Royse C, Chockalingam N. Exploration of implementation, financial and technical considerations within allied health professional (AHP) telehealth consultation guidance: a scoping review including UK AHP professional bodies' guidance. BMJ Open 2021;11(12):e055823 [FREE Full text] [doi: 10.1136/bmjopen-2021-055823] [Medline: 34969656]

Abbreviations

PT: physiotherapist
SLT: speech and language therapist

Edited by C Jacob; submitted 08.07.24; peer-reviewed by A Krithika, T van Houwelingen, L Davies; comments to author 02.12.24; revised version received 22.01.25; accepted 03.03.25; published 26.03.25.

Please cite as:
Rettinger L, Aichinger L, Ertelt-Bach V, Huber A, Javorszky SM, Maul L, Putz P, Sargis S, Werner F, Widhalm K, Kuhn S
Best Practice Guide for Reducing Barriers to Video Call–Based Telehealth: Modified Delphi Study Among Health Care Professionals
JMIR Hum Factors 2025;12:e64079
URL: https://humanfactors.jmir.org/2025/1/e64079
doi:10.2196/64079
PMID:

Exploring User Experience and the Therapeutic Relationship of Short-Term Avatar-Based Psychotherapy: Qualitative Pilot Study

Byeul Jang[1*], MS; Chisung Yuh[2*], DDS, MS; Hyeri Lee[3], MFA; Yu-Bin Shin[1], PhD; Heon-Jeong Lee[1], MD, PhD; Eun Kyoung Kang[4], MD, PhD; Jeongyun Heo[5], PhD; Chul-Hyun Cho[1,6], MD, PhD

[1]Department of Psychiatry, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul, Republic of Korea

[2]Department of Medical Education, Korea University College of Medicine, Seoul, Republic of Korea

[3]Annenberg School for Communication and Journalism, University of Southern California, Los Angeles, CA, United States

[4]KakaoHealthcare Corp., Seongnam, Republic of Korea

[5]Department of Smart Experience Design, Kookmin University, Seoul, Republic of Korea

[6]Department of Biomedical Informatics, Korea University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Chul-Hyun Cho, MD, PhD
Department of Psychiatry, Korea University College of Medicine, 73 Goryeodae-ro, Seongbuk-gu, Seoul, Republic of Korea

Abstract

Background: The rapid advancement of telehealth has led to the emergence of avatar-based psychotherapy (ABP), which combines the benefits of anonymity with nonverbal communication. With the adoption of remote mental health services, understanding the efficacy and user experience of ABP has become increasingly important.

Objective: This study aimed to explore the user experience and therapeutic relationship formation in short-term ABP environments, focusing on psychological effects, user satisfaction, and critical factors for implementation.

Methods: This qualitative study involved 18 adult participants (8 women and 10 men). Participants engaged in two short-term ABP sessions (approximately 50 minutes per session) over 2 weeks, using an ABP metaverse system prototype. Semistructured in-depth interviews were conducted with both the participants and therapists before and after the ABP sessions. The interviews were conducted via an online platform, with each interview lasting approximately 30 minutes. The key topics included the sense of intimacy, communication effectiveness of avatar expressions, emotions toward one's avatar, concentration during sessions, and perceived important aspects of the ABP. Data were analyzed using thematic analysis.

Results: The analysis revealed 3 main themes with 8 subthemes: (1) reduction of psychological barriers through avatar use (subthemes: anonymity, ease of access, self-objectification, and potential for self-disclosure); (2) importance of the avatar–self-connection in therapeutic relationship formation (subthemes: avatar self-relevance and avatar–self-connection fostering intimacy and trust); and (3) importance of nonverbal communication (subthemes: significance of nonverbal expressions and formation of empathy and trust through nonverbal expressions). Participants reported enhanced comfort and self-disclosure owing to the anonymity provided by avatars, while emphasizing the importance of avatar customization and the role of nonverbal cues in facilitating communication and building rapport.

Conclusions: This pilot study provides valuable insights into the short-term ABP user experience and therapeutic relationship formation. Our findings suggest that ABP has the potential to reduce barriers to therapy through anonymity, ease of access, and potential for self-disclosure, while allowing for meaningful nonverbal communication. The avatar–self-connection emerged as a crucial factor in the effectiveness of ABP, highlighting the importance of avatar customization in enhancing user engagement and therapeutic outcomes. Future research and development in ABP should focus on improving avatar customization options, enhancing the fidelity of nonverbal cues, and investigating the long-term effectiveness of ABP compared with traditional face-to-face therapy.

Trial Registration: CRIS KCT0009695; https://tinyurl.com/2a48s7dh

(JMIR Hum Factors 2025;12:e66158) doi:10.2196/66158

KEYWORDS

avatar-based psychotherapy; telehealth; therapeutic relationship; user experience; anonymity; nonverbal communication; mental health; mobile phone

Introduction

Rapid advancements in information technology have led to significant developments in telehealth, particularly in addressing mental health issues, such as anxiety, depression, and stress-related disorders [1]. The COVID-19 pandemic has accelerated the transition from traditional face-to-face clinical services to telehealth solutions [2]. This shift has created new opportunities for individuals with difficulty accessing in-person psychotherapy owing to geographical constraints or personal reservations [3]. Avatar-based psychotherapy (ABP) has gained attention because of its unique ability to combine the benefits of anonymity and nonverbal communication. Recent research has demonstrated the potential effectiveness of metaverse-based counseling approaches compared to traditional in-person settings [4].

The therapeutic relationship between the client and psychotherapist is crucial for successful therapeutic outcomes. However, the dynamics of this relationship in ABP environments require further investigation [5]. Self-disclosure, a key element in relationship building, may be facilitated differently in ABP settings than in traditional face-to-face or video-based online therapy [6]. The concept of avatar–self-connection, which refers to the psychological link between users and their digital representations, may play a significant role in the effectiveness of ABP [7].

This study aimed to explore the user experience and therapeutic relationship formation in short-term ABP environments. Specifically, we sought to understand the psychological effects and satisfaction levels of users engaging in ABP and to identify critical factors to consider when implementing ABP systems.

Methods

Ethical Considerations

This study was conducted in accordance with the principles of the Declaration of Helsinki and approved by the Institutional Review Board of Korea University Anam Hospital (IRB No: 2023AN0504). All participants provided written informed consent before participation. Participants were informed that their participation was voluntary and that they could withdraw at any time without consequences. All interviews were audio-recorded with prior consent and subsequently transcribed for analysis. The data collected in this study were anonymized to ensure the privacy and confidentiality of all participants. This manuscript and its supplementary materials do not include any images or identifiable features of research participants.

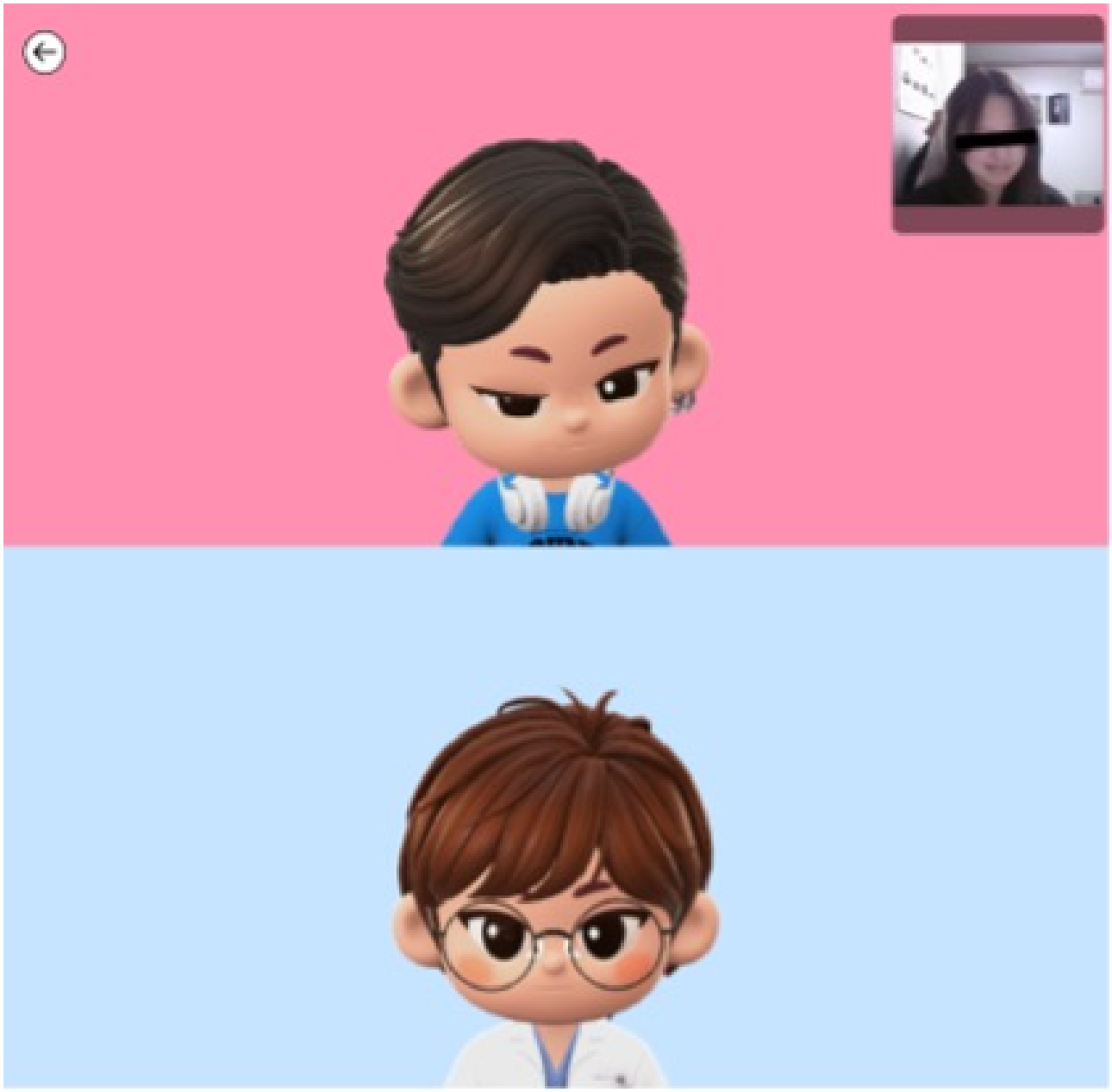
Study Design and Data Collection

This study was designed as a short-term intervention pilot study to explore initial user experiences with ABP. The two-session format was chosen to capture immediate reactions and early therapeutic relationship formation in the ABP environment. We conducted a qualitative study to gain an in-depth understanding of user experiences in ABP sessions. Eligible participants were adults aged 19 - 70 years who self-reported their need for psychological health management and had no previous experience with avatar-based counseling. We excluded individuals with intellectual disabilities, organic brain damage, or those currently receiving psychiatric treatment for major mental disorders (including major mood disorders, anxiety disorders, and schizophrenia spectrum disorders). Additionally, participants were required to have access to and the ability to use smartphones or computers.

An avatar that reflected facial movements in real time was established using a metaverse system prototype (Figure 1). This system allowed participants to engage in psychotherapy sessions through avatars, enabling verbal and nonverbal communication. Eighteen adult participants (8 women) were initially recruited in Seoul, South Korea. The participants engaged in two ABP sessions (approximately 50 minutes per session) over 2 weeks.

Semistructured in-depth interviews were conducted with both the participants and psychotherapists before and after the ABP sessions. These interviews, each lasting 30 minutes, were conducted using the Zoom platform (Zoom Video Communications, Inc.). The key topics included the sense of intimacy and communication effectiveness of avatar expressions, emotions and feelings toward one's avatar, concentration during the session, and perceived important aspects of the ABP sessions.

Figure 1. The avatar-based psychotherapy metaverse system prototype. Top avatar: participant; bottom avatar: psychotherapist. The real face box at the top can be turned off; in a real psychotherapy session, the real face is turned off and is conducted with only the avatar.



Data Analysis

We employed thematic analysis to analyze the data, using ATLAS.ti software (Version 23; ATLAS.ti Scientific Software Development GmbH, Berlin, Germany) [8]. The analysis was conducted in 6 phases: familiarization with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing reports. Four researchers (JB, HL, JH, and CHC) participated in the initial familiarization phase and carefully reviewed the transcripts before individual coding. The 174 initial codes were systematically categorized through iterative analysis. Through team discussions, these codes were refined and consolidated into 3 main themes of psychological barriers (41 codes), avatar–self-connection (29 codes), and nonverbal communication (38 codes). The remaining codes were excluded as they did not directly address the research questions.

Results

The thematic analysis revealed 3 main themes with 8 subthemes, as shown in Table 1. The 174 initial codes were systematically categorized through iterative analysis. Through team discussions, these codes were refined and consolidated into 3 main themes of psychological barriers (41 codes), avatar–self-connection (29 codes), and nonverbal communication (38 codes). The remaining codes were excluded as they did not directly address the research questions.

Table . Thematic framework of user experiences and therapeutic relationship formation in avatar-based psychotherapy: main themes, subthemes, code counts, and representative codes.

Main themes	Subthemes	Code counts	Representative codes
Reduction of psychological barriers through avatar use	1.1 Anonymity	12	• Anonymity reduces social anxiety. • Sense of security through anonymity. • Honest communication enabled. • Anonymity prevents judgment. • Reduced fear of personal recognition.
	1.2 Ease of access	10	• Flexible scheduling. • No travel required. • Accessible from anywhere. • Easier therapy initiation. • Reduced logistical challenges.
	1.3 Self-objectification	8	• Detached view encourages self-reflection. • Avatar mirrors emotional state. • Enhances self-awareness. • Observing oneself via the avatar. • Identifies overlooked emotions.
	1.4 Potential for self-disclosure	11	• Easier to discuss sensitive topics. • Safe space for expression. • Encourages honest sharing. • Enables communication without barriers. • Reduces hesitation in disclosure.
Importance of avatar–self-connection	2.1 Avatar self-relevance	15	• Personalized avatars foster comfort. • Realism enhances connection. • Reflects personal identity. • Improves immersion. • Mirrors user's traits effectively.
	2.2 Fostering intimacy and trust	14	• Facilitates faster rapport. • Acts as a conversational bridge. • Builds therapeutic trust. • Supports meaningful dialogue. • Encourages emotional openness.

Main themes	Subthemes	Code counts	Representative codes
Importance of nonverbal communication	3.1 Significance of nonverbal expressions	20	• Gestures enhance communication. • Nonverbal cues reduce misunderstandings. • Expressive avatars improve immersion. • Adds emotional depth. • Reinforces communication effectiveness.
	3.2 Empathy and trust formation	18	• Mirrors emotions effectively. • Conveys empathy through gestures. • Builds trust via nonverbal cues. • Enhances emotional understanding. • Reduces virtual communication gaps.

Theme 1: Reduction of Psychological Barriers Through Avatar Use

Theme 1, the reduction of psychological barriers through avatar use, emerged as the most prominent theme with 41 codes. These codes were distributed across 4 subthemes: anonymity (12 codes), ease of access (10 codes), self-objectification (8 codes), and potential for self-disclosure (11 codes). Participants reported that the use of avatars in psychotherapy sessions helped lower psychological barriers primarily through 3 mechanisms.

Subtheme 1.1: Anonymity

The anonymity provided by avatars was frequently mentioned as a key factor in facilitating open and honest communication. One participant noted, “Knowing that the therapist does not know my real face gives me a sense of security. Even if we met in real life, they would not recognize me, which is reassuring in a way.”

Subtheme 1.2: Ease of Access

The convenience of accessing ABP services through a metaverse platform was highlighted as a significant advantage. One participant stated, “The fact that psychotherapy sessions can be done from anywhere makes it more accessible and easier to form a rapport. I think anonymous ABP is more effective in this regard.”

Subtheme 1.3: Self-Objectification

Avatars facilitated self-objectification, allowing participants to view themselves from a detached perspective. One participant observed, “When I occasionally looked at my avatar, I noticed that I was not smiling much. It made me think, ‘I look a bit pitiful.’ This showed me how I appeared.”

Subtheme 1.4: Potential for Self-Disclosure

Avatars facilitated self-disclosure. One participant noted, “For people who want psychotherapy without revealing themselves too much, expressing themselves through an avatar in the metaverse can make it easier to talk honestly with the therapist.”

Theme 2: Importance of Avatar–Self-Connection in Therapeutic Relationship Formation

The importance of avatar–self-connection in therapeutic relationship formation emerged as the second major theme, comprising 29 codes. These were distributed between 2 key subthemes: avatar self-relevance (15 codes) and avatar–self-connection fostering intimacy and trust (14 codes). The results showed that the degree of connection between users and their avatar played a crucial role in the therapeutic relationship formation within the ABP environment.

Subtheme 2.1: Avatar Self-Relevance

Participants expressed a desire for their avatars to reflect aspects of their real selves. One participant noted, “If I could customize the avatar more to my liking and feel that it truly represents me, I think I would feel more completely connected to it.”

Subtheme 2.2: Avatar–Self-Connection Fostering Intimacy and Trust

A strong avatar–self-connection enhanced the sense of intimacy and trust in the therapeutic relationship. One participant shared, “When I felt that this avatar was really me, I could immerse myself more deeply in the session and form a stronger therapeutic relationship with the therapist.”

Theme 3: Importance of Nonverbal Communication

The importance of nonverbal communication emerged as a significant theme with 38 codes, distributed across 2 subthemes: significance of nonverbal expressions (20 codes) and formation of empathy and trust through nonverbal expressions (18 codes). The thematic analysis highlighted the significant role of nonverbal communication in ABP, emphasizing its impact on self-expression, empathy, and trust-building.

Subtheme 3.1: Significance of Nonverbal Expressions

The participants emphasized the importance of nonverbal cues in enhancing communication. One participant stated, “If the avatar could show hand gestures or body language, I think it would be even better. These physical movements can have psychological significance.”

Subtheme 3.2: Formation of Empathy and Trust Through Nonverbal Expressions

Nonverbal communication through avatars enhanced empathy and trust in the therapeutic relationship. One participant shared, "I think the avatar helped by expressing my gestures and expressions to some extent. This somewhat overcomes communication errors that can occur when we are not face-to-face."

Discussion

Study Findings and Comparison With Previous Findings

This study provides critical insights into ABP, highlighting 3 interconnected dimensions: psychological barrier reduction, avatar–self-connection, and the role of nonverbal communication in metaverse therapeutic systems.

The anonymity, accessibility, and potential for self-disclosure provided by avatars create a unique therapeutic environment that lowers the resistance to mental health engagement, consistent with previous research on online self-disclosure [5]. A novel finding of this study was the self-objectification through avatar use, suggesting a potential mechanism for enhanced self-awareness that resonates with self-distancing theories in psychology [9]. This indicates that ABP may offer a natural pathway for adaptive self-reflection, a key component in many therapeutic approaches.

The self-objectification observed in this study demonstrates an interesting contrast with findings from video conferencing research, which found that viewing one's actual image during video conferences can lead to greater cognitive burden and negative psychological effects, particularly for individuals with high public self-consciousness [10]. However, our findings suggest that avatar-mediated self-observation may offer distinct advantages. Unlike video conferencing where users see their actual image, avatar representation creates a beneficial psychological distance that facilitates more objective self-reflection while maintaining emotional engagement. This unique characteristic of avatar-mediated interaction may explain why participants in our study reported enhanced self-awareness without the negative psychological impacts often associated with direct self-viewing in video conferences. The avatar-mediated environment provides a unique balance between self-awareness and psychological comfort, potentially making it particularly suitable for therapeutic contexts where self-reflection is crucial but emotional safety needs to be maintained.

Furthermore, the avatar–self-connection played a crucial role in the therapeutic relationship formation, extending previous work on avatar realism [7]. Our findings directly linked this connection to the quality of the therapeutic relationship, underscoring the potential significance of avatar customization in fostering engagement and improving outcomes. The avatar may serve as a bridge between the client's inner world and the therapeutic space, facilitating deeper exploration of personal narratives.

Our observations of nonverbal communication in ABP expand on previous research, highlighting how even limited nonverbal cues can significantly impact the therapeutic process [11]. This adaptability suggests that ABP taps into fundamental aspects of human communication, offering a rich alternative to traditional face-to-face therapy.

The interplay among these elements in ABP creates a unique therapeutic ecosystem that balances self-disclosure and self-protection. As we continue to explore the potential of ABP, it is crucial to consider its place within the broader context of mental health care. ABP can be viewed as a complementary tool that can lead to more comprehensive and accessible care, potentially reaching individuals who might otherwise not receive treatment [12,13].

Strengths and Limitations

The findings of this study have several implications for clinical practice. First, ABP may be particularly beneficial for clients hesitant to engage in traditional face-to-face therapy because of social anxiety, stigma, or geographical limitations [14]. The anonymity and ease of access provided by ABP could serve as a stepping stone for these individuals to transition to other forms of therapy, if needed. Second, clinicians using ABP should be aware of the importance of the avatar–self-connection and consider ways to enhance this connection during therapy sessions. This might involve discussing the client's choice of avatar and exploring how it relates to their self-perception or therapeutic goals. However, although ABP offers many advantages, clinicians must be aware of its limitations. The reduced nonverbal cues in ABP compared with face-to-face therapy require clinicians to develop new skills in reading and interpreting client expressions and emotions through avatars.

Conclusion

These findings not only contribute to the growing body of knowledge on telehealth interventions, but also prompt us to reconsider the nature of the therapeutic relationship in the digital age. As technology continues to reshape our modes of interaction, ABP stands at the forefront of a new paradigm in mental health care that harnesses the power of the virtual world to foster real-world healing and growth.

Acknowledgments

This work was supported by grants from the National Research Foundation (NRF) of Korea (NRF-2021R1A5A8032895 and NRF-2022M3C1B6080866), a grant from the Institute of Information and Communications Technology Planning and Evaluation (IITP; RS-2023-00224823), and a grant from the Information and Communications Promotion Fund (H0601-24-1017) and Broadcasting Communications Development Fund (D1013-24-1003) through the National IT Industry Promotion Agency (NIPA),

Republic of Korea. The funders had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the paper for publication.

Authors' Contributions

JH and CHC are cocorresponding authors and are responsible for the data and materials, manuscript submission, peer review, publication process, authorship details, and ethics committee approval. EKK, JH, and CHC conceived and designed the study. JB, CY, HL, JH, and CHC performed data analyses. JB, CY, HL, JH, and CHC wrote the first draft of the manuscript. JB, HL, EKK, JH, and CHC participated in data collection. All authors edited all versions of the manuscript. All authors were involved in interpreting the results and have read, commented on, and approved the final version of the manuscript.

Conflicts of Interest

EKK is a member of Kakao Healthcare, the company that developed the avatar-based psychotherapy metaverse system prototype used in this study. The other authors declare no conflicts of interest.

References

1. Williams MG, Stott R, Bromwich N, Oblak SK, Espie CA, Rose JB. Determinants of and barriers to adoption of digital therapeutics for mental health at scale in the NHS. BMJ Innov 2020 Jul;6(3):92-98. [doi: 10.1136/bmjinnov-2019-000384]
2. Reay R, Kisely SR, Looi JCL. Better Access: substantial shift to telehealth for allied mental health services during COVID-19 in Australia. Aust Health Rev 2021 Dec;45(6):675-682. [doi: 10.1071/AH21162] [Medline: 34743787]
3. Braude G, Mohi S, Quinlan E, Shoullis A, Collison J. Psychologists and their self-mirror: therapy experiences via videoconferencing. Clin Psychol (Aust Psychol Soc) 2023 Sep 2;27(3):384-391. [doi: 10.1080/13284207.2023.2245953]
4. Cho S, Kang J, Baek WH, Jeong YB, Lee S, Lee SM. Comparing counseling outcome for college students: Metaverse and in-person approaches. Psychother Res 2024 Nov;34(8):1117-1130. [doi: 10.1080/10503307.2023.2270139] [Medline: 37848177]
5. Bente G, Rüggenberg S, Krämer NC, Eschenburg F. Avatar-mediated networking: increasing social presence and interpersonal trust in net-based collaborations. Human Comm Res 2008 Apr;34(2):287-318. [doi: 10.1111/j.1468-2958.2008.00322.x]
6. Hooi R, Cho H. Avatar-driven self-disclosure: The virtual me is the actual me. Comput Human Behav 2014 Oct;39:20-28. [doi: 10.1016/j.chb.2014.06.019]
7. Kim DY, Lee HK, Chung K. Avatar-mediated experience in the metaverse: The impact of avatar realism on user-avatar relationship. J Retail Consum Serv 2023 Jul;73:103382. [doi: 10.1016/j.jretconser.2023.103382]
8. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
9. Kross E, Ayduk O. Making meaning out of negative experiences by self-distancing. Curr Dir Psychol Sci 2011 Jun;20(3):187-191. [doi: 10.1177/0963721411408883]
10. Kuhn KM. The constant mirror: Self-view and attitudes to virtual meetings. Comput Human Behav 2022 Mar;128:107110. [doi: 10.1016/j.chb.2021.107110]
11. Bailenson JN, Yee N, Merget D, Schroeder R. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. Pres Teleoperators Virtual Environ 2006 Aug 1;15(4):359-372. [doi: 10.1162/pres.15.4.359]
12. Kim SW, Baek SH, Cho CH. Psychosocial and digital intervention for patients with schizophrenia. J Korean Med Assoc 2024 Feb;67(2):103-114. [doi: 10.5124/jkma.2024.67.2.103]
13. Sestino A, D'Angelo A. My doctor is an avatar! The effect of anthropomorphism and emotional receptivity on individuals' intention to use digital-based healthcare services. Technol Forecast Soc Change 2023 Jun;191:122505. [doi: 10.1016/j.techfore.2023.122505]
14. Cho CH. The beginning of the era of digital therapeutics in Korea: challenges and opportunities. J Korean Med Sci 2023 May 29;38(21):e166. [doi: 10.3346/jkms.2023.38.e166] [Medline: 37270919]

Abbreviations

ABP: avatar-based psychotherapy

Edited by A Kushniruk; submitted 05.09.24; peer-reviewed by D Jung, H Kwon; revised version received 02.12.24; accepted 23.12.24; published 05.02.25.

Please cite as:
Jang B, Yuh C, Lee H, Shin YB, Lee HJ, Kang EK, Heo J, Cho CH
Exploring User Experience and the Therapeutic Relationship of Short-Term Avatar-Based Psychotherapy: Qualitative Pilot Study
JMIR Hum Factors 2025;12:e66158
URL: https://humanfactors.jmir.org/2025/1/e66158
doi:10.2196/66158

Original Paper

The Research Agenda for Perinatal Innovation and Digital Health Project: Human-Centered Approach to Multipartner Research Agenda Codevelopment

Haneen Amhaz[1,2,3], BSc, MPH; Sally Xuanping Chen[1,2], BSc, MFS; Amanee Elchehimi[1,2], MA, MPH; Kylin Jialin Han[1,2], BSc, BE, MFS, MBA; Jade Morales Gil[1,2], BSN; Lu Yao[1,2], BA; Marianne Vidler[1,4], MPH, PhD; Kathryn Berry-Einarson[5], BA, MPH; Kathryn Dewar[1], PhD; May Tuason[6,7], BSN, MBA; Nicole Prestley[1], BA; Quynh Doan[2,8], MHSc, MDCM, FRCPC, PhD; Tibor van Rooij[2,9], BSc, PhD; Tina Costa[6], BSN, MA; Gina Ogilvie[1,2,10], MSc, DrPH, FCFP, MD; Beth A Payne[1,2,10], BSc, PhD

[1]Women's Health Research Institute, Vancouver, BC, Canada

[2]BC Children's Hospital Research Institute, Vancouver, BC, Canada

[3]Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada

[4]Department of Obstetrics and Gynaecology, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

[5]Quality and Research, Perinatal Services BC, Provincial Health Services Authority, Vancouver, BC, Canada

[6]Provincial Digital Health and Information Services, Provincial Health Services Authority, Vancouver, BC, Canada

[7]BC Ministry of Health, Victoria, BC, Canada

[8]Department of Pediatrics, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

[9]Department of Computer Science, Faculty of Science, The University of British Columbia, Vancouver, BC, Canada

[10]School of Population and Public Health, Faculty of Medicine, The University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Beth A Payne, BSc, PhD
Women's Health Research Institute
H214-4500 Oak Street, Box 42
Vancouver, BC, V6H3N1
Canada
Phone: 1 604 875 3459
Email: bpayne@cw.bc.ca

Abstract

Background: Digital health innovations provide an opportunity to improve access to care, information, and quality of care during the perinatal period, a critical period of health for mothers and infants. However, research to develop perinatal digital health solutions needs to be informed by actual patient and health system needs in order to optimize implementation, adoption, and sustainability.

Objective: Our aim was to co-design a research agenda with defined research priorities that reflected health system realities and patient needs.

Methods: Co-design of the research agenda involved a series of activities: (1) review of the provincial Digital Health Strategy and Maternity Services Strategy to identify relevant health system priorities, (2) anonymous survey targeting perinatal care providers to ascertain their current use and perceived need for digital tools, (3) engagement meetings using human-centered design methods with multilingual patients who are currently or recently pregnant to understand their health experiences and needs, and (4) a workshop that brought together patients and other project partners to prioritize identified challenges and opportunities for perinatal digital health in a set of research questions. These questions were grouped into themes using a deductive analysis approach starting with current BC Digital Health Strategy guiding principles.

Results: Between September 15, 2022, and August 31, 2023, we engaged with more than 150 perinatal health care providers, researchers, and health system stakeholders and a patient advisory group of women who were recently pregnant to understand the perceived needs and priorities for digital innovation in perinatal care in British Columbia, Canada. As a combined group, partners were able to define 12 priority research questions in 3 themes. The themes prioritized are digital innovation for (1) patient

autonomy and support, (2) standardized educational resources for patients and providers, and (3) improved access to health information.

Conclusions: Our research agenda highlights the needs for perinatal digital health research to support improvements in the quality of care in British Columbia. By using a human-centered design approach, we were able to co-design research priorities that are meaningful to patients and health system stakeholders. The identified priority research questions are merely a stepping stone in the research process and now need to be actioned by research teams and health systems partners.

(JMIR Hum Factors 2025;12:e60825) doi:10.2196/60825

KEYWORDS

digital health; co-design; digital strategy; human-centered design; eHealth; cocreation; codevelopment; perinatal intervention; quality of care; digital tools; pregnancy; patient autonomy; patient support; mobile phone

Introduction

The digital health sector continues to expand after being an integral part of health care delivery during the COVID-19 pandemic, resulting in a growing interest in digital health research and innovation. Digital health provides an opportunity to reduce health disparities by addressing barriers and facilitators to health care access and improving the health and lifestyle of those who use it [1]. Digital innovation is also essential to offering value-based health care, which involves changes in health systems that are measured by improvement in patient outcomes per unit health care expenditure [2]. Core to achieving improvements in value-based care is establishing an understanding of what value is based on identifying outcomes that matter to patients themselves.

Human-centered design (HCD) is an approach to problem-solving that can be used to increase value in health care by focusing on problems and outcomes of interest to patients. HCD also has a role in improving health equity [3,4]. Core characteristics of an HCD approach include an iterative process of empathy and discovery to identify the core problem and then systematic brainstorming of solutions where the human experiencing the problem plays a central role [5]. The HCD approach goes beyond simply working with the end user to focus on their interaction with the technology and considers the broader social context in which the technology is used and other partners who may indirectly be impacted by the use of the technology [4,5]. The result of working in partnership with people under these assumptions using the HCD approach is meant to be a product or service that meets real human needs and reflects their values. This approach has been used to solve complex health care challenges, such as improving pediatric patient experience in acute care [6] and increasing adherence to self-care practices for chronic conditions [7,8]. In a perinatal health setting, these methods have been used to develop digital interventions for gestational diabetes [9].

The perinatal period encompasses a person's experiences from conception to the first year after delivery including pregnancies that end in miscarriage or stillbirths. The majority of patients receiving perinatal care in British Columbia has access to a mobile device and the internet and are increasingly looking to digital technologies to inform and support their care [10,11]. Using digital technology in health care can improve the quality of care and health outcomes for patients and infants during this early stage of life. The use of digital health tools in perinatal care has already been shown to result in lower and healthier gestational weight, lower smoking rates, increased physical activity, and lower rates of maternal and infant morbidity (eg, fewer preterm births, gestational diabetes, and pre-eclampsia), leading to fewer planned and unplanned visits to health care facilities [1,12-14]. Digital health has also been demonstrated to reduce the practitioners' workload as well as relieve some of the resource strains on an overburdened system [15]. Due to the vast potentiality of digital health and opportunities for further innovation, it is important to engage with patients who have received perinatal care and care providers as end users to identify emerging priorities for future innovations and research that are informed by actual needs [16]. By identifying and responding to real patient and care provider needs, we can optimize the implementation, adoption, and sustainability of perinatal digital health solutions, create value, and ensure existing health inequities are not exacerbated [16,17].

In this paper, we report on the Research Agenda for Perinatal Innovation and Digital Health (RAPID) project [18], which was led by the BC Children's Hospital Research Institute and Women's Health Research Institute in partnership with the Provincial Health Services Authority (PHSA) and Perinatal Services BC (PSBC) with the aim of codeveloping research priorities with patients and care providers.

Methods

Ethical Considerations

This project was run as a knowledge translation initiative, and no research ethics board approval was required. The engagement activities conducted with partners were classified as quality improvement activities rather than systematic investigations to advance knowledge. Based on these distinctions, our institutional research ethics board determined that the project falls outside of the scope of research subject to ethics review, and a waiver was granted. All workshop and survey participants consented to the anonymous collection of their opinions. Members of our patient advisory committee were compensated hourly according to local standards to reflect their equal standing to research team members within the project.

Project Design and Setting

The patient and health system partners with whom we worked were partners in the research priority setting through co-design and were compensated for their time. All final workshop

participants provided consent to the collection of their opinions and views during the workshop and through the follow-up survey.

The objectives of the RAPID project were to facilitate activities that bring key partners and existing knowledge together to (1) identify gaps in the perinatal continuum of care that can be addressed with digital innovation to improve patient and family health and (2) define current barriers to optimal adoption of digital health interventions for perinatal health care from the patient and health system perspectives. This knowledge was then used to define research priorities and codevelop research questions to motivate digital innovation informed by care needs and addressing barriers.

Recruitment of Partners and Governance

Following guidance outlined by Concannon et al [19], we developed an engagement plan for our activities that describes an analysis of which partner groups should be involved, their roles and modes of participation, and sampling strategies (ie, snowball sampling). Partners to include in engagement activities were grouped as (1) people with lived experience of perinatal care who were pregnant or recently delivered a baby; (2) perinatal health care providers including nurses, midwives, physicians, and other allied health; and (3) health care decision makers who are involved with the design or delivery of perinatal health care in British Columbia.

This project was governed by two primary groups: (1) a steering committee and (2) a patient advisory committee (PAC). Using snowball sampling, starting with a core group of experts at BC Children Hospital Research Institute and Women's Health Research Institute, we identified key partners with expertise in research, innovation, and implementation processes related to perinatal, maternal, or digital health services to form the steering committee. The core group was involved from the initiation of the project including during the development of the funding proposal. This core group included 4 health system decision makers as research users and 4 researchers. Additional partners identified for the steering committee included 2 clinical care providers and 2 additional researchers. The final steering committee included 5 researchers, 3 health care providers, and 4 health system decision makers. The purpose of the steering committee was to oversee project objectives and align efforts to each of their organizations' known digital health priorities and ongoing work.

Patient partners for the advisory committee were recruited through social media and the local hospital patient engagement office. The criteria for participation were (1) multilingual and (2) pregnant or a caregiver of a child younger than 2 years of age. There were 15 respondents to the recruitment advertisements, 9 of whom responded to our invitation to complete a 15-minute informal interview so we could learn more about their background, and they could also learn more about our project. After the interviews, the steering committee was able to select 4 participants who were willing and able to commit the time required by the project and who had diverse perinatal experiences (based on type of care provider and mode of delivery), geographic location in British Columbia, and backgrounds (language spoken and profession). The PAC met separately with only 4 members of the study team and steering committee during the initial activities of the project. This was done to create a safe space that would limit the impact of known health system hierarchies and preserve the authentic patient voice in the identification of perinatal health system pain points and the development of initial insights related to these pain points [20]. Once initial insights were defined with the patient advisors, all partners were brought together in a final workshop to collaborate and build consensus on research priorities as described in more detail below.

Engagement Activities

A series of engagement activities with each distinct group of partners was completed before a final workshop that brought all partners and results together to reach a consensus on final research priorities.

Perinatal Care providers

Perinatal care providers across British Columbia including physicians, midwives, nurses, and allied health were invited to share, rate, and explore their thoughts on digital perinatal health using a ThoughtExchange (version 6.4; Fulcrum Management Solutions Ltd). ThoughtExchange is a web-based survey tool that allows participants to share their thoughts in response to a prompt, which can be anonymously viewed and ranked by other participants. The platform then creates weighted rankings of the thoughts on the prompted topic to show participant priorities. The care providers were asked to respond to the following prompt: "What digital health tools are most valuable to your delivery of perinatal health care currently, and what do you think are the greatest areas of opportunity for digital innovation going forward?" The survey was promoted through all British Columbia health authority internal staff newsletters and through member email lists at Doctors of BC; Family Doctors of BC; and University of British Columbia Faculty of Medicine alumni, School of Midwifery, Departments of Obstetrics and Gynaecology, Pediatrics and Family Medicine. Three rounds of promotion were completed. Once the advertised deadline passed, the highest-ranked thoughts were reviewed by 2 research members of the steering committee (BAP and SXC), and key messages were developed through an inductive approach. Basic descriptive statistics were used to analyze the ranking of thoughts.

Health System Partners

Steering committee members from the British Columbia Ministry of Health (MoH), the PHSA, and the PSBC were each asked to review their organizations' recently published reports outlining their strategies for addressing current gaps in health. The reports reviewed were the BC Digital Health Strategy (BCDHS) [21] and the PSBC Maternity Services Strategy (MSS). Access to the MSS was provided through confidential personal communication with PSBC Director Robert Finch. These reports were presented at a steering committee meeting within the first 3 months of the project to identify priorities and alignments specific to perinatal digital health within these documents. The steering committee members from each organization outlined ongoing digital innovation projects in perinatal care and priorities specific to digital health solutions

for improving perinatal care. Each report, postpresentation, was summarized in lay terms to highlight priorities. These summaries were reviewed and edited by the health system partners to ensure that summaries were comprehensive, not leaving out key information.

Patients

A series of four 90-minute PAC meetings were held and recorded over Zoom (Zoom Video Communications). They served to ascertain current digital health use and perceived need for new innovation as well as to inform workshop development from a patient perspective. The meetings included a series of HCD and user-experience research activities, participatory discussions, and postmeeting homework that typically took an hour to complete. The first meeting served as an introduction to all members; the following 2 meetings were developed to allow patients to share their experiences and understand how they interact with the health care system in order to identify gaps, common touchpoints with health care systems and services, and needs (Table 1). For the final meeting, the PAC was provided the lay summaries from the engagement with other partners prior to the meeting. They were asked to review these and share their reflections during the meeting. In addition, they were asked to develop insight statements that identify a need or gap. All meetings were facilitated by the core research team consisting of BAP, SXC, HA, and MV, who were also steering committee members. Activities were carried out on Miro (version 2.0; Miro), a web-based collaborative workspace. Meeting notes were collected and circulated to all participants for review and input. After each meeting, study team members also recorded reflective notes.

Table 1. Patient engagement meeting plan.

Session number	Objective	Presession activity	Activities
1	Establish an environment where patients feel safe and comfortable sharing their experiences	• N/A[a]	• Ice breaker • Introduction to digital health • Empathy mapping activity
2	Gather patients' perspectives and experiences in the health system during pregnancy and after birth	• Complete an empathy map [22] and a 30-minute reflection on their experiences in perinatal care	• Share empathy map and experiences • Introduce user journey activity
3	Identify health system touchpoints and potential areas with opportunities for improvement	• Complete a user journey [23] • Reflect on fellow participants' empathy maps and reflections	• Open discussion regarding user journey and experiences • Introduce insightful statements
4	Prepare for workshops with patients	• Review summary of ThoughtExchange and health systems' report summaries • Insightful statements worksheet [24]	• Comment on summary results • Go over insightful statements • Plan workshop

[a]N/A: not applicable.

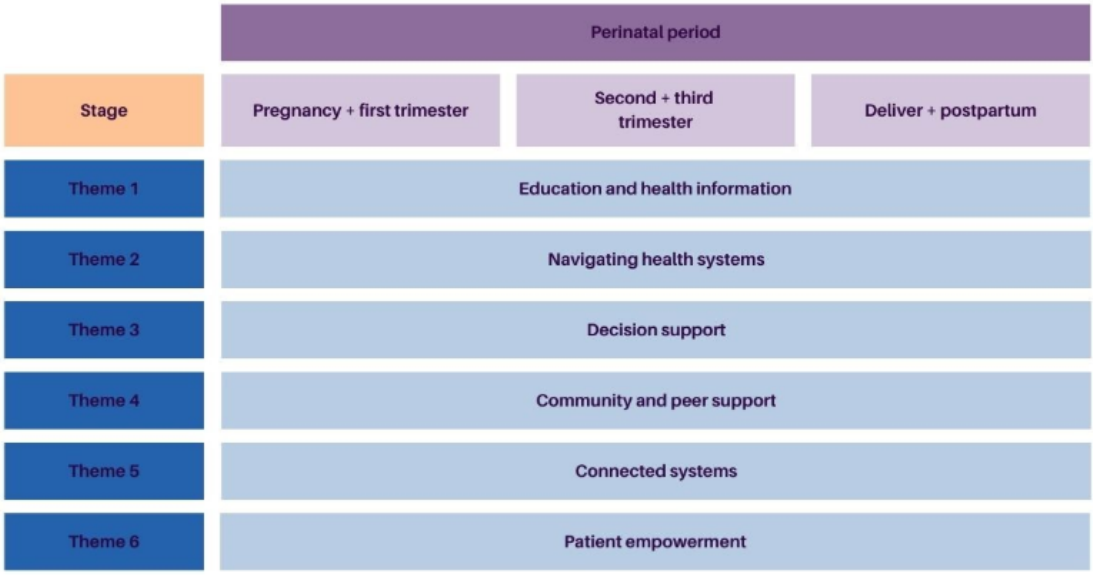
Final Consensus Workshop and Analysis

A final 90-minute workshop was held to bring together all the partners and interested parties to share and align on challenges and priorities collected from earlier engagement sessions. This workshop was composed of both existing members of each partner group: patients, clinical care providers, researchers and health system decision makers as well as additional participants recruited through snowball sampling.

In preparation for the workshop, the core study team (BAP, SXC, HA, and MV) reviewed insight statements from the patients, and using a combination of deductive (using the guiding principles of the BCDHS as a starting point) and inductive analysis, placed these into overarching and emerging themes and then further grouped statements based on relevance to a stage of the perinatal period as described in Figure 1. This grouping was done to make the review and brainstorming session more feasible to complete within our limited 90-minute time frame. Research representatives from the steering committee, BAP and MV, developed 2 example research questions for each thematic area to prime the workshop discussion. These questions were developed to prompt discussion during the workshop, considering many participants do not have a background in research.

Attendees were divided into the 3 perinatal stage–based groups based on their expertise. For example, an attendee who is a neonatologist was assigned to the delivery+postpartum group. Each group contained at least 1 member from each of the partner types in order to facilitate collaboration across groups and consensus building. Each group was asked to reflect on and revise the provided insight statements and example research questions as well as develop their own research questions in order to identify research priorities based on their knowledge and experience. At the end of the session, each group was asked to identify 3-5 priority research questions from their brainstorming exercise. Each group then presented their priority research questions to the full workshop, and collectively, all attendees further discussed and ranked the research questions in order of priority.

Figure 1. Workshop groups and themes.



After the workshop, the research team reviewed meeting recordings, notes, and research question rankings to finalize a list of priority research questions that reflected workshop consensus. These research questions were then reviewed a final time by the PAC to ensure patient needs were at the forefront. This final list of research questions was converted and deployed as a REDCap (Research Electronic Data Capture; Vanderbilt University) survey to all workshop attendees. Respondents were asked to select 3 research questions they believed to be of the highest priority for perinatal care. Survey rankings were summarized descriptively to identify priorities from within the defined research questions.

Results

Engagement activities occurred between September 15, 2022, and August 31, 2023. The steering committee met monthly during this time to review progress and plan adjustments to the strategy, as new knowledge was developed.

Thought Exchange With Clinical Care Providers

A total of 125 perinatal health care provider participants responded, 102 of whom shared what provincial health authority they were affiliated with. British Columbia is split by geographic regions into 5 regional health authorities with an additional province-wide authority focused on Indigenous health and the PHSA that provides oversight for all regions. Most participants came from PHSA (n=25, 24.5%), Fraser Health (n=19, 18.6%), Vancouver Coastal Health (n=16, 15.7%), or Island Health (n=16, 15.7%). There were no participants from the First Nations Health Authority, and 7 participants did not affiliate with any of British Columbia's 7 health authorities.

Of the 125 participants, 55 responded with 80 different thoughts; 51 participants rated the thoughts of their colleagues, resulting in a total of 1108 ratings. Analysis of the thoughts based on the built-in functions of the ThoughtExchange platform identified the highest-ranked themes, which were grouped and summarized. Five key takeaways were identified under the themes of health education and connected systems from the ThoughtExchange:

1. There is an urgent need to create more inclusive, responsive, and accessible digital health tools. The end user (patient or health care provider) must be involved in the design of these tools to make sure that they are useful and can solve the right problems.
2. Among all current digital health tools shared, "Zoom," which is a tool that allows videos and remote health visits, and the "Hyperbilirubinemia risk assessment tool," which is a way for doctors to check if a baby is at risk of developing a condition where their skin and eyes turn yellow, are the highest ranked. "Up to Date," a web-based platform that keeps clinicians up-to-date with the latest medical research and guidelines (the United States only), was mentioned but ranked lower. A tool like "Up to Date" that provides the latest guidelines specific to Canada was identified as a current need to support high-quality care.
3. Important clinical areas that were identified as being priority topics for digital innovation were postpartum lactation support and mental health services. There is a demand for innovation in these areas to be accessible to individuals residing in rural regions.
4. A 2-way communication system is needed to support sharing health information across the province and between patients and providers; this means patients should have equal access to their electronic medical records as their health care providers. Patients should also be able to freely share their information between various levels of health care units at their own discretion.
5. Provincially standardized, free, evidence-based, and user-friendly prenatal patient education was identified as a need that could be solved with digital innovation. Accessible prenatal education would support high-quality perinatal care for all patients in British Columbia.

Health System Partner Engagement

A review of the MoH and PSBC digital health reports and strategies with health system partners in the steering committee meetings identified opportunities where perinatal innovation could be used to improve the quality of the health care system in British Columbia.

The MoH report, BCDHS, identified the following gaps in care across British Columbia: lack of coordination between health care services, poor communication systems, and inadequate follow-up. Therefore, their digital health priorities include (1) empowering patients through participation in their health journey (eg, inclusive, trusted, and equitable health content provided digitally), (2) increasing the capacity for providers to deliver high-quality care (eg, improving digital literacy and promoting the use of digital tools), (3) establishing a connected health system that allows for the sharing of data (eg, interoperability across health authorities and provider level and collaborating on clinical solutions), and (4) infrastructure and business processes streamlined for efficiency (eg, modernizing British Columbia's health supply chain, data flow, asset tracking, and analytics). To accommodate these strategies and goals, PSBC is working on developing a digital platform that acts as a central place for patient information. Based on the identified gaps, strategies, and needed tools, three potential areas for research were identified: (1) What is the best way to use telecommunication and information technologies to provide health care services during pregnancy and childbirth remotely? (2) What is the best design and implementation strategy to enable a shared electronic antenatal health record that can be accessed by any health care provider and patient during pregnancy? (3) Can we validate the use of wearable devices that can track and monitor health data during pregnancy and childbirth (ie, remote blood pressure for hypertension in pregnancy and blood glucose monitoring for gestational diabetes)?

PSBC's MSS report focuses particularly on addressing modifiable factors such as maternal weight and nutrition, substance use, mental health, diabetes, infant sleep, and breastfeeding. They currently have a series of projects in progress, and those specific to digital health solutions include a personalizable information and tool resource hub for providers and patients, a single source of data that can be shared across multiple systems, leveraging access to clinical data for quality improvement, allowing for enhanced digital training opportunities, increasing access to telemedicine care, and updated screening protocols and quality control. Based on these approaches five areas for future research were identified: (1) With the implementation of new technologies, what gaps and biases arise that impact their inclusivity and equity? (2) Although we can assess the immediate and short-term benefits of these technologies, how do we assess their long-term impact on maternal and child health? (3) How do we improve interoperability between different digital platforms to enable seamless data sharing and communication? (4) What policies would need to be put in place to ensure that patient data and privacy are protected? How would this integrate patient consent? (5) How do we understand the needs of the user (patients and health care service providers) and design resources that are user-centered? How do we ensure that users will engage with these resources?

Patient Engagement Sessions

The 4 members of the PAC varied in languages spoken (1 Mandarin, 1 Cantonese, 1 Spanish, and 1 Arabic), age of their children (1 still pregnant at the start of the project, 1 younger than 6 months, and 2 between 6 months and 1 year), mode of delivery (2 vaginal birth and 2 cesarean section), and location of their birth experience (2 Vancouver Coastal Health, 1 Fraser Health, and 1 Interior Health). Over the course of our 4 group meetings, we completed empathy maps, user journeys (Multimedia Appendix 1), and insight statements and collected reflective notes and recordings of live discussions. Common challenges were experienced among all patient partners navigating the perinatal period. For example, all patient partners described difficulties in understanding the different types of perinatal care providers and how to find an available midwife or doctor in their region. They all also spoke of challenges in navigating the transition from pregnancy to early newborn care, particularly when they had nonurgent concerns about their new baby because many available health care resources focus on urgent and emergency conditions.

In the final PAC meeting, we reflected on the summaries of the ThoughtExchange, MSS report, and BCDHS to identify recurring themes, priorities, and gaps that were also expressed by patients, which were then presented in a master list of insight statements (Textbox 1). These insights were divided into 6 themes: education or health information, navigating health systems, decision support, community or peer support, connected systems, and patient empowerment. Overall, over the 4 meetings, patients clearly expressed that they wanted digital innovation that would support their confidence and ability to navigate the health system in all regions of the province of British Columbia where health care is currently distributed and often fragmented. This should come in the form of free and accessible educational resources and tools to navigate the perinatal health journey. All patient partners discussed how stressful it is to navigate the current system and the frustration of having to communicate their health information to different care providers repeatedly. Almost all of our patient partners moved during the course of their pregnancy and experienced challenges finding maternity care providers both initially and after they moved. Another universal experience was the challenge presented by our current system, which does not allow for the pairing of the mother's and infant's health records. This often resulted in logistical challenges in navigating postpartum and postnatal follow-up that were borne entirely by the new parents. As a result of the success of these sessions, the research team worked with the PAC to develop a toolkit for patient engagement that was made available for use by other research teams (Multimedia Appendix 2).

Textbox 1. Final list of insight statements.

Reliable health information is needed

- Patients who are accessing perinatal care are overwhelmed by navigating an abundance of resources, and they are not sure which ones are accurate to follow.
- When seeking health-related information on the internet, information is often conflicting or misleading.
- When seeking health information on the internet, information is only in English and can be difficult to understand or communicate to multilingual family members.
- Health information provided is in technical terms that patients are unfamiliar with leaving patients confused.
- Patients often rely on informal peer support, which adds to the conflicting information they receive, leaving them more confused and unsure of what information to trust.
- Due to the abundance of information available on the internet, patients feel like they are wasting or spending too much of their time filtering through the information.
- Patients indicated that it is challenging to understand the difference between an obstetrician and a midwife especially for newcomers or immigrants from a different health care system; they do not know how to choose the right professional for them.

Lack of supports

- Due to the duration between pregnancy and a patient's first obstetrician gynecologist appointment being too long, patients are left with the responsibility of seeking information independently.
- There are limited perinatal classes that are poorly advertised and are not free; therefore, not all patients have the means to access them.
- Patients are navigating aspects of the health care system during their perinatal journey that they have never had to use before; this often comes with little guidance and doubts about what steps need to be taken.
- Patients who are pregnant for the first time do not know where and what services to seek, leaving them feeling helpless and guilty if complications occur.

Need for continuum of care

- Patients' care is interrupted by the transfer of care to a different care provider especially after delivery. These periods of interruption leave patients uncertain and lost.
- After being discharged after delivery or from midwives, patients feel like their only support is the emergency room, especially if they do not have a family doctor.

Mental health challenges

- Complications during the perinatal period leave patients feeling incompetent and highly anxious about their own situations.
- Patients wanting to seek mental health support are discouraged because they are unaware of how to access them or told they are not in need of them.

Emergency screening

- Due to the long wait times in the emergency room and lack of knowledge of problems arising either during prepartum or postpartum, patients indicated a need to have an emergency screening process tailored to pregnant and newborn mothers by using the digital platform before they head to the hospitals to save or rule out unnecessary visits.

Uncoordinated transitions of care

- It was challenging for patients to find an appropriate obstetrician or midwife in a timely manner; they had to call a list of clinics on their own or find one through word of mouth in order to be accepted.
- When patients find out they are pregnant, they are frustrated by a lack of information on available care providers in their region.
- Patients are confused about when to choose a midwife or other type of doctor and why they may need to switch providers during pregnancy.
- Mom and baby's care is not coordinated in the health system after the birth of the baby because they are treated as 2 distinct patients.
- There is no standardized system to share health records between perinatal care providers, which leaves patients with the responsibility to communicate their health information when multiple types of caregivers are needed to support their pregnancy.
- When accessing care, services, or resources across different health authorities, health authorities do not follow standardized structures or guidelines acting as an additional barrier for patients as they learn to navigate a new health authority.
- There is a lack of communication and awareness among health authorities, therefore, interrupting patients' care.

Poor patient-centered care

- Care providers are dismissive of patient concerns and thoughts, resulting in patients advocating for themselves and leaving them stressed and isolated.
- Patients feel uninvolved in their health because they are often told what to do without direct and clear explanations or guidelines from their providers.
- Patients have to manage multiple appointments with different providers instead of an integrated care team, meaning that they often wait long periods of time before their needs are addressed.
- Patients would appreciate more visits and opportunities to meet with their care providers or a trusted source in order to discuss their concerns and ask questions.
- Patients want to hear follow-up from their care provider after every test, even if it is normal.

Final Workshop

The workshop had a total of 20 attendees: 4 core project team members, 4 patient partners, 2 researchers, 5 clinicians, and 5 health system partners. The workshop was held over Zoom. During the workshop, 3 research themes were chosen as priorities through the discussion process including patient autonomy and support, educational resources for patients and providers, and access to health information (Figure 2). Attendees chose to frame the final research questions around broad transformative goals rather than specific clinical care concepts or processes to make the final agenda broadly applicable across health settings in the British Columbia public health system. For example, although patients and clinicians identified a need for lactation support earlier in the engagement process, a decision was made not to name this clinical concern specifically, as it was felt to narrow the scope of the recommendations. Similarly, patients and care providers identified a need for more accessible support for nonurgent health concerns. Suggestions were made to create more digital resources, but in discussion, the group agreed that more foundational knowledge on the type of digital technology that would be found most usable and effective for providing health information is needed. This resulted in final research questions such as 1.3 and 2.3. A total of 12 final research questions to be included in the research agenda were developed to reflect the identified priorities and patient perspectives (Figure 3).

Figure 2. Three priority themes for perinatal digital health research.

Figure 3. Final list of priority research questions. RPM: remote patient monitoring.

A follow-up REDCap survey was sent to a total of 35 individuals, including all workshop participants, steering committee members, and patient partners, and we received 20 responses. Respondents included 5 researchers, 5 health system decision makers, and 6 perinatal health care providers with 4 patient advisors. When asked to select the top 3 priority research questions, respondents prioritized questions 1.4, 1.2, and 3.3 (Textbox 2). All of these questions focus on supporting patients' decision-making with better information and fall within the themes of improving patient autonomy and access to information. When comparing the priority of the overarching themes, overall questions from theme 1: patient autonomy and support were the most selected; however, if we adjust for the number of questions per category, questions from theme 3: access to health information were the most selected.

Textbox 2. Prioritization of research questions.

Number of times prioritized: 9

- Question 1.4: What aspects of perinatal care can be monitored, administered, or assessed using mobile devices (ie, remote patient monitoring, digital health visits, and peer support) to support empowered self-care? How do we implement these without widening the access to care gap experienced by already disadvantaged populations?

Number of times prioritized: 7

- Question 1.2: What digital tools and information can support patients' decisions in identifying and connecting with the right maternity care provider? What information do patients need to identify the right maternity care provider? How do we support this decision with digital tools?
- Question 3.3: How can the integration of patient-reported outcomes in a patient's record contribute to quality improvement and shared decision-making?

Number of times prioritized: 5

- Question 1.5: How do we monitor the mental health patients receiving perinatal care in order to provide timely support when needed and improve health outcomes and perinatal recovery?
- Question 2.1: What resources can be compiled into a starter kit to guide patients through their pregnancy? What resources can be compiled into a starter kit to guide patients who have just delivered their child? How do these resources impact patient and child health outcomes?
- Question 2.3: How do patients prefer to view and access health information and educational resources (eg, type of format such as audio vs video vs website)? How can we engage patients and providers in developing and designing these resources? Are there existing resources we can expand in function to meet patient needs (eg, 811 response, peer support, chatbots, or SmartMom)?
- Question 3.1: What is the effect of more transparent access to provincial dashboards such as HealthGateway (for both parent and child and linked together) on care quality, and access to care in both urban and rural or remote settings?

Number of times prioritized: 4

- Question 3.2: Can improved data access for patients and care providers lead to an increase in British Columbia's status against global benchmarks for health care quality?

Number of times prioritized: 3

- Question 2.2: What are the individual, organizational, and policy barriers and facilitators to providing standardized multilingual prenatal and postnatal education in British Columbia?

Number of times prioritized: 2

- Question 1.1: How do we best measure patient autonomy in the perinatal period?
- Question 1.3: What services and resources exist within British Columbia that support patient mental health, self-care, and patient autonomy in decision-making? How do we ensure providers are aware of them? How do we emphasize connecting patients to them?

Number of times prioritized: 1

- Question 2.4 What implementation strategies will work to make educational resources accessible across regions in British Columbia? How do we measure impact?

Discussion

Research Priorities and Implications for Practice

The RAPID project resulted in the co-design of a comprehensive research agenda for perinatal digital innovation that reflects provincial strategies and patient experience in British Columbia. During the project, a series of engagement activities allowed us to ascertain the current gaps and needs for digital health interventions for perinatal care and develop a research agenda that will address these needs. Our research agenda provides a foundation for identifying, developing, and implementing digital solutions that are responsive and feasible within the current British Columbia health care system landscape. The RAPID research agenda identified a core set of research priorities under 3 interdependent thematic areas that, if addressed, could lead to meaningful improvements in patient and clinician experience and health outcomes in the perinatal period. The priority themes identified are patient autonomy and support when navigating the health system, patient and provider access to education resources, and patients' ability to access health information. These themes align with the BCDHS and have broad support from health system leaders.

The research themes identified and research questions prioritized in this project suggest that further innovation is needed to properly leverage digital health tools to support perinatal care in British Columbia. Priority research questions and a major area of discussion with our patient partners focused on a need for better digital tools to support self-care and decision-making across the entire perinatal period, from pregnancy to the end of the first year of the infant's life. The ThoughtExchange also highlighted that, within the British Columbian health system, there is a limited number of tools in use that clinicians view as

valuable to care. This results in patients accessing numerous external digital resources that are not always of quality and relevance to their care. This agenda can be used to advocate for change at the provincial level and support the incorporation of digital health into the health system. Addressing research priorities identified by RAPID would improve the patient experience by providing needed solutions that allow patients to safely interact with their health care professionals through web-based platforms or through their mobile devices, access self-care tools and applications, use smartphones or other devices to manage and monitor their health, and adopt technology in their health plans.

Digital health innovation as suggested by our research agenda is appealing because it is not restricted by time or place [13] allowing for patient empowerment as well as patient participation in the decision-making of their own health. Addressing the RAPID research priorities can also improve the continuity of care and delivery of essential services, something that all partners in the RAPID project would clearly value. There is already evidence that digital innovation in perinatal care can lead to improved health outcomes for both the woman or pregnant person and their infant [1,12-14]. This agenda highlights that the perceived value of digital innovations for patients receiving perinatal care and their families cannot only be determined by traditional health outcomes. Digital innovations should also be measured against their ability to improve confidence and autonomy in decision-making, self-care, information sharing, and patient and clinical experience.

The digital health sector receives enormous investment, making available a vast number of solutions varying in quality and capabilities [17]. The existing literature has provided a broad understanding of the feasibility, acceptability, and effectiveness of digital tools outlining advantages and disadvantages [1,12]. However, this continued growth in digital solutions does not guarantee or equate to an increase in solutions that can be implemented to meet the actual existing needs of patients, providers, and health systems [17]. In this project, we used HCD methods to identify patient needs and priorities for digital innovation that are aligned with clinical care providers and health system priorities, but our identified priorities cannot exist in isolation from the current knowledge of socioecological factors. Although the RAPID priority research questions are defined based on need, these must be explored with consideration for the facilitators and barriers that contribute to engagement with digital health solutions. Previous research on digital health use has identified barriers and facilitators to implementation such as infrastructure, equipment, network access, motivation, and usability [25,26].

The effort was made to integrate the knowledge of barriers and facilitators to engagement and uptake of digital health innovations throughout the research agenda, such as geographic settings, data-security concerns, and diversity in health concerns. However, considerations for these barriers and facilitators should continue beyond what the questions have defined, as we work to action the agenda. Issues such as connectivity, user motivation, social support, digital literacy, user-friendly design, and credibility should continue to be examined throughout subsequent digital health research.

Strengths and Limitations

This project's use of a human-centered approach prioritized patient voices and provided the space for collaborative co-design of the final research questions. Limiting the number of patient partners and restricting the number of researchers in the patient advisory meetings allowed for better quality engagement through intimate and extensive discussions. The diversity of partners from across British Columbia allows for more generalizable results. However, a significant limitation was our lack of engagement with Indigenous clinical care providers or patients. There is a need to engage with Indigenous people's voices within all of the partner groups before moving research forward from this strategy. Related to this, although we were able to collect a diverse group of partners' opinions and bring these together to design a comprehensive research agenda, running the project under a knowledge translation initiative restricted the amount of data we could collect on engaged participants. As a result, it may be more difficult for others to apply this knowledge outside of the British Columbia health system. We also cannot explore if participant characteristics such as age or experience level (for health care providers) had any impact on their opinions of digital health priorities. An additional strength is our engagement with health system partners throughout the project, aligning in strategy and garnering early support from decision makers. As a result, outcomes of research informed by this agenda are more likely to be implemented and achieve long-term impact. There was strong collaboration throughout the project between researchers and knowledge users, allowing us to lead this work within the health system, leveraging developed partnerships, and articulating alignment with organizational goals and patient research priorities.

This HCD approach prioritized patient voices in developing research priorities; yet, our multipartner codevelopment approach recognizes that the ownness should not be on patients alone to identify a solution to the complex issues of perinatal health care delivery. The emphasis on people solving their own problems that is core to the HCD approach cannot be directly assumed in a health care context due to the complexity of the system and the multitude of interested parties that will be impacted by any system changes. In positioning our work to address patient needs while still aligning with health system priorities, we have aimed to support value-based care efforts in the British Columbia health system. Focusing further research and innovation efforts to address the RAPID research priorities can support improvements in value-based care, as they would improve outcomes and processes that matter to patients without an unreasonable increase in health care expenditure.

Conclusions

This project identified clear support and enthusiasm across partners for digital innovation to improve the quality of perinatal care in the province of British Columbia. Our research agenda highlights the need for better patient education and health information systems that are more accessible to patients and care providers with the overarching need for digital resources that can increase patient autonomy in decision-making throughout pregnancy and the transition to newborn care. Co-design of this research agenda is a first step in the research

process. Our priority now is to take action on identified research priorities. Action can occur through traditional research funding pathways but also by advocating to our health system partners to directly address needs identified by the RAPID research agenda within the health system.

Acknowledgments

The authors gratefully acknowledge that our project could not have been possible without the invaluable contributions of our patient partners AE, KJH, JMG, and LY. The authors thank Joanne Hepplewhite from the Provincial Health Services Authority Transformation development office for her support with the ThoughtExchange. This project was funded through a convening and collaborating grant (#C2-2022-2969) from the Michael Smith Health Research BC.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Anonymized example of a user journey.
[PNG File , 2366 KB - humanfactors_v12i1e60825_app1.png]

Multimedia Appendix 2
Toolkit for patient engagement.
[PDF File (Adobe PDF File), 1612 KB - humanfactors_v12i1e60825_app2.pdf]

References

1. Lim S, Tan A, Madden S, Hill B. Health professionals' and postpartum women's perspectives on digital health interventions for lifestyle management in the postpartum period: a systematic review of qualitative studies. Front Endocrinol (Lausanne) 2019;10:767 [FREE Full text] [doi: 10.3389/fendo.2019.00767] [Medline: 31787932]
2. Teisberg E, Wallace S, O'Hara S. Defining and implementing value-based health care: a strategic framework. Acad Med 2020;95(5):682-685 [FREE Full text] [doi: 10.1097/ACM.0000000000003122] [Medline: 31833857]
3. Holeman I, Kane D. Human-centered design for global health equity. Inf Technol Dev 2019;26(3):477-505 [FREE Full text] [doi: 10.1080/02681102.2019.1667289] [Medline: 32982007]
4. Evans L, Evans J, Pagliari C, Källander K. Scoping review: exploring the equity impact of current digital health design practices. Oxf Open Digit Health 2023;1:oqad006. [doi: 10.1093/oodh/oqad006]
5. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. Int J Qual Health Care 2021;33(Suppl 1):37-44 [FREE Full text] [doi: 10.1093/intqhc/mzaa127] [Medline: 33068104]
6. Simons R. Improving the patient experience for children at the paediatric acute medical unit. Delft University of Technology. 2018. URL: https://resolver.tudelft.nl/uuid:7a975ce4-1529-4065-bcdb-b323cf0747a0 [accessed 2018-04-19]
7. Matheson G, Pacione C, Shultz RK, Klügl M. Leveraging human-centered design in chronic disease prevention. Am J Prev Med 2015;48(4):472-179. [doi: 10.1016/j.amepre.2014.10.014] [Medline: 25700655]
8. Wildevuur S, Thomese F, Ferguson J, Klink A. Information and communication technologies to support chronic disease self-management: preconditions for enhancing the partnership in person-centered care. J Particip Med 2017;9(1):e14. [doi: 10.2196/jopm.8846] [Medline: 36262005]
9. Yee LM, Leziak K, Jackson J, Strohbach A, Saber R, Niznik CM, et al. Patient and provider perspectives on a novel mobile health intervention for low-income pregnant women with gestational or type 2 diabetes mellitus. J Diabetes Sci Technol 2021;15(5):1121-1133 [FREE Full text] [doi: 10.1177/1932296820937347] [Medline: 32627582]
10. So long landline, hello smartphone. Statistics Canada. 2023. URL: https://www.statcan.gc.ca/o1/en/plus/3582-so-long-landline-hello-smartphone [accessed 2023-04-09]
11. Peragallo Urrutia R, Berger AA, Ivins AA, Beckham AJ, Thorp JM, Nicholson WK. Internet use and access among pregnant women via computer and mobile phone: implications for delivery of perinatal care. JMIR Mhealth Uhealth 2015;3(1):e25 [FREE Full text] [doi: 10.2196/mhealth.3347] [Medline: 25835744]
12. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
13. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile health for perinatal depression and anxiety: scoping review. J Med Internet Res 2020;22(4):e17011 [FREE Full text] [doi: 10.2196/17011] [Medline: 32281939]
14. Hearn L, Miller M, Fletcher A. Online healthy lifestyle support in the perinatal period: what do women want and do they use it? Aust J Prim Health 2013;19(4):313-318. [doi: 10.1071/PY13039] [Medline: 23899373]

15. Pham Q, El-Dassouki N, Lohani R, Jebanesan A, Young K. The future of virtual care for older ethnic adults beyond the COVID-19 pandemic. J Med Internet Res 2022;24(1):e29876 [FREE Full text] [doi: 10.2196/29876] [Medline: 34994707]
16. Figueroa CA, Luo T, Aguilera A, Lyles CR. The need for feminist intersectionality in digital health. Lancet Digit Health 2021;3(8):e526-e533 [FREE Full text] [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
17. Mathews SC, McShea MJ, Hanley CL, Ravitz A, Labrique AB, Cohen AB. Digital health: a path to validation. NPJ Digit Med 2019;2:38 [FREE Full text] [doi: 10.1038/s41746-019-0111-3] [Medline: 31304384]
18. RAPID: Research Agenda for Perinatal Innovation & Digital Health. Women's Health Research Institute. URL: https://whri.org/our-initiatives/rapid/ [accessed 2023-05-23]
19. Concannon TW, Grant S, Welch V, Petkovic J, Selby J, Crowe S, et al. Practical guidance for involving stakeholders in health research. J Gen Intern Med 2019;34(3):458-463 [FREE Full text] [doi: 10.1007/s11606-018-4738-6] [Medline: 30565151]
20. Trauma-informed practice guide. BC Provincial Mental Health and Substance Use Planning Council. 2013. URL: https://cewh.ca/wp-content/uploads/2012/05/2013_TIP-Guide.pdf [accessed 2024-11-12]
21. B.C.'s Digital Health Strategy. 2023. URL: https://divisionsbc.ca/sites/default/files/72475/MoH-Digital-Health-Strategy-Handout-Feb-2023.pdf [accessed 2024-12-12]
22. Build your creative confidence: empathy maps. IDEO. URL: https://www.ideo.com/journal/build-your-creative-confidence-empathy-maps [accessed 2024-04-22]
23. Build your creative confidence: customer journey map. IDEO. URL: https://www.ideo.com/journal/build-your-creative-confidence-customer-journey-map [accessed 2024-04-22]
24. The field guide to human-centered design. IDEO. 2015. URL: https://www.designkit.org/resources/1.html [accessed 2024-12-12]
25. Kaboré SS, Ngangue P, Soubeiga D, Barro A, Pilabré AH, Bationo N, et al. Barriers and facilitators for the sustainability of digital health interventions in low and middle-income countries: a systematic review. Front Digit Health 2022;4:1014375 [FREE Full text] [doi: 10.3389/fdgth.2022.1014375] [Medline: 36518563]
26. Moulaei K, Moulaei R, Bahaadinbeigy K. Barriers and facilitators of using health information technologies by women: a scoping review. BMC Medical Informatics and Decision Making 2023;23(1):176 [FREE Full text] [doi: 10.1186/s12911-023-02280-7] [Medline: 37670281]

Abbreviations

BCDHS: BC Digital Health Strategy
HCD: human-centered design
MoH: Ministry of Health
MSS: Maternal Services Strategy
PAC: patient advisory committee
PHSA: Provincial Health Services Authority
PSBC: Perinatal Services BC
RAPID: Research Agenda for Perinatal Innovation and Digital Health
REDCap: Research Electronic Data Capture

Edited by C Jacob; submitted 23.05.24; peer-reviewed by H Witteman, MG Onaiwu, E Kjörk; comments to author 23.08.24; revised version received 15.10.24; accepted 26.11.24; published 30.01.25.

Please cite as:
Amhaz H, Chen SX, Elchehimi A, Han KJ, Gil JM, Yao L, Vidler M, Berry-Einarson K, Dewar K, Tuason M, Prestley N, Doan Q, van Rooij T, Costa T, Ogilvie G, Payne BA
The Research Agenda for Perinatal Innovation and Digital Health Project: Human-Centered Approach to Multipartner Research Agenda Codevelopment
JMIR Hum Factors 2025;12:e60825
URL: https://humanfactors.jmir.org/2025/1/e60825
doi:10.2196/60825
PMID:

Original Paper

Building a Decentralized Biobanking App for Research Transparency and Patient Engagement: Participatory Design Study

Ananya Dewan[1], BS, BA; M Eifler[2], MFA; Amelia Hood[3], MA; William Sanchez[2], BS; Marielle Gross[2,3], MBE, MD

[1]Johns Hopkins University School of Medicine, Baltimore, MD, United States

[2]de-bi, co., Baltimore, MD, United States

[3]Johns Hopkins Berman Institute of Bioethics, Baltimore, MD, United States

Corresponding Author:
Marielle Gross, MBE, MD
Johns Hopkins Berman Institute of Bioethics
1809 Ashland Avenue
Baltimore, MD, 21205
United States
Phone: 1 8135416103
Email: mariellesophiagross@gmail.com

Abstract

Background: Patient-derived biospecimens are invaluable tools in biomedical research. Currently, there are no mechanisms for patients to follow along and learn about the uses of their donated samples. Incorporating patients as stakeholders and meaningfully engaging them in biomedical research first requires transparency of research activities.

Objective: In this paper, we describe the use of participatory design methods to build a decentralized biobanking "de-bi" mobile app where patients could learn about biobanking, track their specimens, and engage with ongoing research via patient-friendly interfaces overlaying institutional biobank databases, initially developed for a breast cancer use case.

Methods: This research occurred in 2 phases. In phase 1, we designed app screens from which patients could learn about ongoing research involving their samples. We embedded these screens in a survey (n=94) to gauge patients' interests regarding types of feedback and engagement opportunities; survey responses were probed during 6 comprehensive follow-up interviews. We then held an immersive participatory design workshop where participants (approximately 50) provided general feedback about our approach, with an embedded codesign workshop where a subset (n=15) provided targeted feedback on screen designs. For phase 2, we refined user interfaces and developed a functional app prototype in consultation with institutional stakeholders to ensure regulatory compliance, workflow compatibility, and composability with local data architectures. We presented the app at a second workshop, where participants (n=25, across 9 groups) shared thoughts on the app's usability and design. In this phase, we conducted cognitive walkthroughs (n=13) to gain in-depth feedback on in-app task navigation.

Results: Most of the survey participants (61/81, 75%) were interested in learning the outcomes of research on their specimens, and 49% (41/83) were interested in connecting with others with the same diagnosis. Participants (47/60, 78%) expressed strong interest in receiving patient-friendly summaries of scientific information from scientists using their biospecimens. The first design workshop identified confusion in terminology and data presentation (eg, 9/15, 60% of co-designers were unclear on the biospecimens "in use"), though many appreciated the ability to view their personal biospecimens (7/15, 47%), and most were excited about connecting with others (12/15, 80%). In the second workshop, all groups found the app's information valuable. Moreover, 44% (5/9) noted they did not like the onboarding process, which was echoed in cognitive walkthroughs. Walkthroughs further confirmed interest in biospecimen tracking, and 23% (3/13) had confusion about not finding any of their biospecimens in the app. These findings guided refinements in onboarding, design, and user experience.

Conclusions: Designing a patient-facing app that displays information about biobanked specimens can facilitate greater transparency and engagement in biomedical research. Co-designing the app with patient stakeholders confirmed interest in learning about biospecimens and related research, improved presentation of data, and ensured usability of the app in preparation for a pilot study.

(JMIR Hum Factors 2025;12:e59485) doi:10.2196/59485

KEYWORDS

mobile health; mHealth; application; smartphone; digital health; digital intervention; participatory design; biobanking; research transparency; donation; patient-derived biospecimens; plain language communications; patient education

Introduction

Background

Samples of blood, tissue, urine, and other biological products, collected from patients during a medical procedure, are invaluable tools in biomedical research. Reviews estimate that >600 research biobanks exist in the United States, housing tens of millions of samples from tens of millions of patients [1,2]. Biospecimens are used in a vast array of foundational, translational, and clinical research [3-6]. Research on biospecimens shows exceptional promise in studying cancer and developing new cancer treatments [4,7,8]. Cutting-edge specimen preservation methods offer high-fidelity disease models for genomic analysis, high-throughput drug testing, advanced imaging, and more.

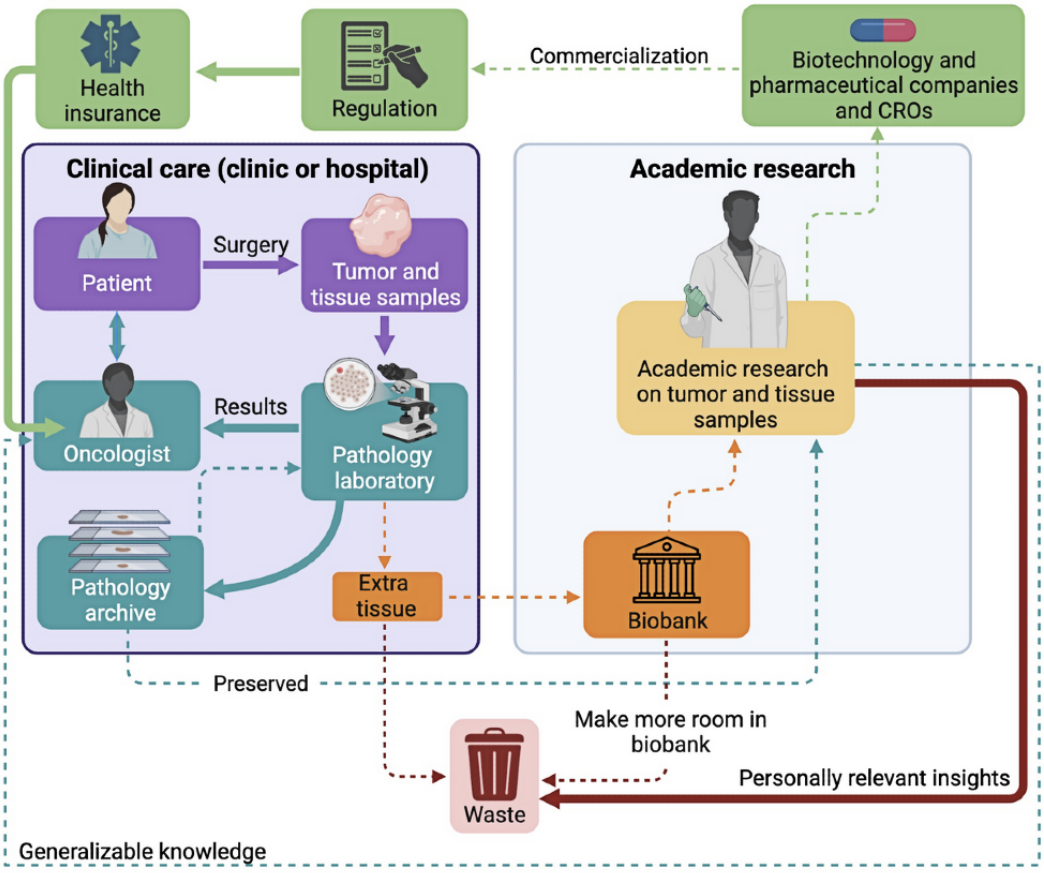
Typically, leftover surgical samples are collected after a 1-time informed consent process per the federal Common Rule in the United States, where patients receive information about the proposed research [9]. The terms of this consent include no expectation to follow up with pertinent research results after the form is signed and samples are collected [10]. In the transition from clinical by-product to banked research sample, patient identifiers are removed from biospecimen data [11]. Deidentification is meant to protect patient privacy by removing identifiable information; however, its conditions preclude scientists from sharing follow-up information about if, when, and how they use biospecimens for research [12]. Scientists receive deidentified biospecimens from the biobank and have no mechanism to directly communicate personally relevant insights or engage with the people whose samples they have received (Figure 1) [10]. Thus, while clinical results from pathology reports are communicated to patients, information on how their biospecimens are used and associated research findings are never shared.

Incorporating patients as stakeholders and meaningfully engaging them in biomedical research is increasingly recognized as critical to more ethical and effective scientific advancement [13]. Patients' lived experiences can help guide research questions, priorities, and methodologies, ensuring that the resulting knowledge aligns with their and their community's priorities. Various innovative technologies and techniques have enhanced patient engagement in research [14,15]; for instance, the James Lind Alliance method facilitates collaboration among clinicians, the public, and additional stakeholders, enabling their opinions to converge and inform the establishment of research priorities [16,17]. Digital platforms that assemble citizens from different backgrounds to share experiences and actively codevelop innovative health strategies have been proposed [18]. However, improving transparency surrounding the nature and extent of research activities is a prerequisite to achieving true patient engagement.

Decentralizing biobank information and empowering biospecimen donors with knowledge about how their samples are used lays the foundation for fostering trust and transparency in biospecimen research [19]. This is particularly vital in rebuilding trust with communities considered marginalized, which have historically been subject to exploitation and unethical research practices [20]. Building trust lays the foundation for further participation by diverse communities and ensures that research endeavors are accountable to those whose biospecimens make the research possible [21].

In previous work, we have demonstrated that there is unmet patient demand for increased transparency surrounding biospecimen research use [22]. We surveyed patients with breast cancer (n=109) about their interest in learning about biobank research and their individual biospecimens [22]. This survey measured patient interest in a broad range of information that they could theoretically learn about research on their samples. Notably, survey respondents were interested in learning information that is readily available within in local biobank and research databases, such as whether their samples had been used in research (66%), details about research conducted on their samples (53%), and the ability to see images of their biospecimens if available (68%).

Figure 1. Biobanking ecosystem: biospecimen supply chain from clinical by-products to institutionalized assets. In the current biobanking paradigm, personally relevant insights and extra tissue are often wasted, representing lost opportunity to advance patient care and maximize value from donated biospecimens. CRO: contract research organization.



Objectives

To address the unmet need for greater transparency and engagement between biobanks and patients, we sought to build an initial decentralized biobanking app—"de-bi"—to serve as a patient's "window" into an established biobank database. We aimed to build a composable platform that would be accessible and valuable to patients while being compliant with existing regulations, composable with data architectures and compatible with established workflows [23]. By viewing biobank data, patients can be empowered to "track" what happens to their biospecimens [24]. A patient-facing mobile app could overlay and display existing data, serving as a technological tool for increasing the transparency of research activities and as a mechanism for directly engaging with patients who have contributed to research, all within the current frameworks of consent and deidentification. Composability with institutional databases would also serve as a source of efficiency, easing the burden of the initial manual data transfer to an app interface and setting up easier future automation and scalability. This tool could then be piloted to measure its impact on patient participation, understanding, and attitudes about biobanking and biospecimen research and begin to answer outstanding empirical questions about effectively fulfilling obligations of transparency and engagement.

Including patients in the design phase was imperative to the future pilot's success. Before the tool was deployed, we sought to ensure that it was interesting, accessible, understandable, valuable, and easy and ideally pleasant to use. Here, we describe the process and results of using participatory design methods to define, design, and deliver a mobile app that enables patients to learn about their biospecimens collected for research, the status of the biospecimens' use in research protocols, and overall biobank collections, as demonstrated in a breast cancer biobank setting.

Methods

Overview

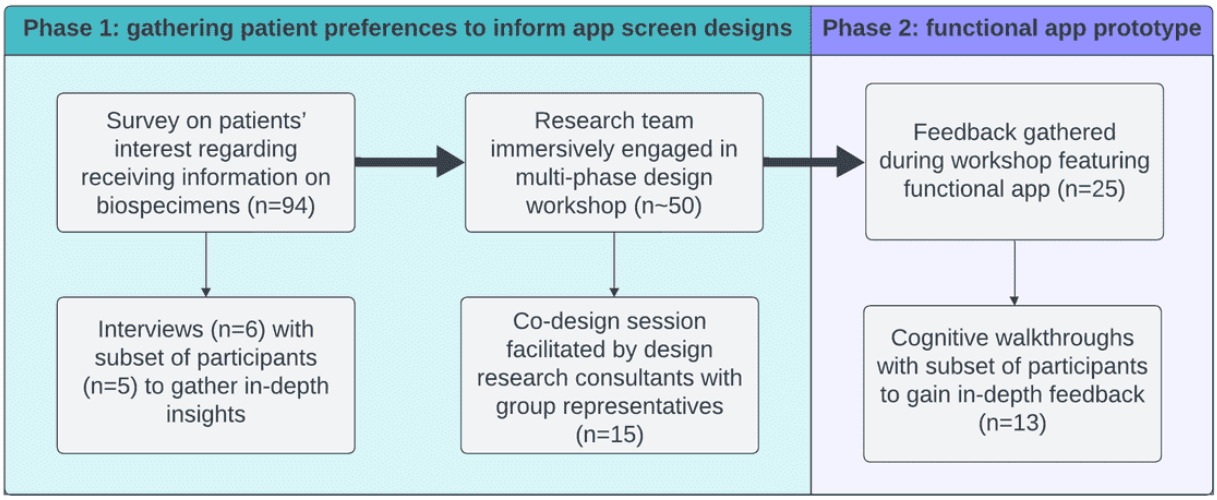
This multiphase participatory design study was conducted among patients of a breast cancer clinic at a large US academic institution. It consisted of surveys of patients (n=94), six semistructured in-depth follow-up interviews with five survey respondents (5/94, 5%), immersive design workshops with patients (workshop 1: approximately 50 participants, focusing on a subset of 15 codesign participants), workshop 2: n=25), and cognitive walkthroughs with individual patients (n=13) [25]. The data being presented represent a subset of broader participant engagement, including additional workshops and encounters with the local community of patients with breast cancer.

This research occurred in 2 phases (Figure 2). In phase 1, we designed smartphone screens containing different information that patients could learn about biobank and research activities

involving their samples. Embedding these screen designs in a survey, we sought to gauge patients' interest in receiving information about research or their biospecimens. We engaged a subset of survey respondents in short web-based interviews to discern their views on the importance of having this information and their opinions on its presentation and design. We held a design workshop at which participants gave feedback on the screens and suggested improvements.

For phase 2, we refined the user interfaces and developed a functional app prototype. As we developed the app, we consulted institutional stakeholders to enhance compatibility with regulations and composability of core app features, with one another and with local data architectures. We then distributed the app at a second design workshop, where participants shared thoughts on the usability and design of the app. In this phase, we conducted cognitive walkthroughs with individual participants to observe their success in using the app and to gain in-depth feedback on its functionality [26].

Figure 2. Overview of study methods.



Phase 1

App Screen Design

We used Whimsical and Figma for the initial wireframes and Flutter and Adalo for prototyping to develop the initial screen designs [27,28]. The content of these designs was informed by previous engagement with this patient population and limited by the content available within institutional biobank or research databases and publicly accessible research content. Previous survey results identified important data points to include in our patient-facing display and informed its design [22]. Working with investigators on the biobanking research protocol, their data managers, and the University of Pittsburgh Institutional Review Board, we selected data points that would be feasible and acceptable to return to patients in a future, functional solution. We then created a sample screenshot presenting a model set of samples and the corresponding hypothetical data points.

We designed 3 initial app screens presenting information about a hypothetical user's collection of biospecimens in the biobank (Figure 3). These screens were meant to be viewed sequentially, with increasing specificity of information presented about individual biospecimens in the biobank. The interface was inspired by the term "biobanking" and designed to resemble many banking apps where users first view a summary of their accounts and can click through to increase specificity about each one. Our first screen contained a summary list of a hypothetical individual's inventory of donated samples, including pictures of each. The second screen displayed high-level demographic information about each individual biospecimen: an image, its location, and the date it was collected. The third and final screen showed more details about each individual sample, such as its ID number, the medical procedure from which it was collected, the number of the freezer where it was stored, and more.

A fourth screen was created for potential users to learn about the biobank collection as a whole (Figure 4). This screen, called "connect," showed a list of subcommunities of biobank donors sorted into groups by their diagnosis code, including summary statistics indicating how many donors to the biobank shared a diagnosis (eg, invasive ductal carcinoma). The design suggests that users might be members of ≥1 of these groups (called "labs") and might use their membership to "connect" with other users, patients, or scientists.

Figure 3. The 3 "biowallet" screens, as developed in Adalo and featured in the survey, presenting different categories of specimens using accordion unfolding and individual sample–level details.

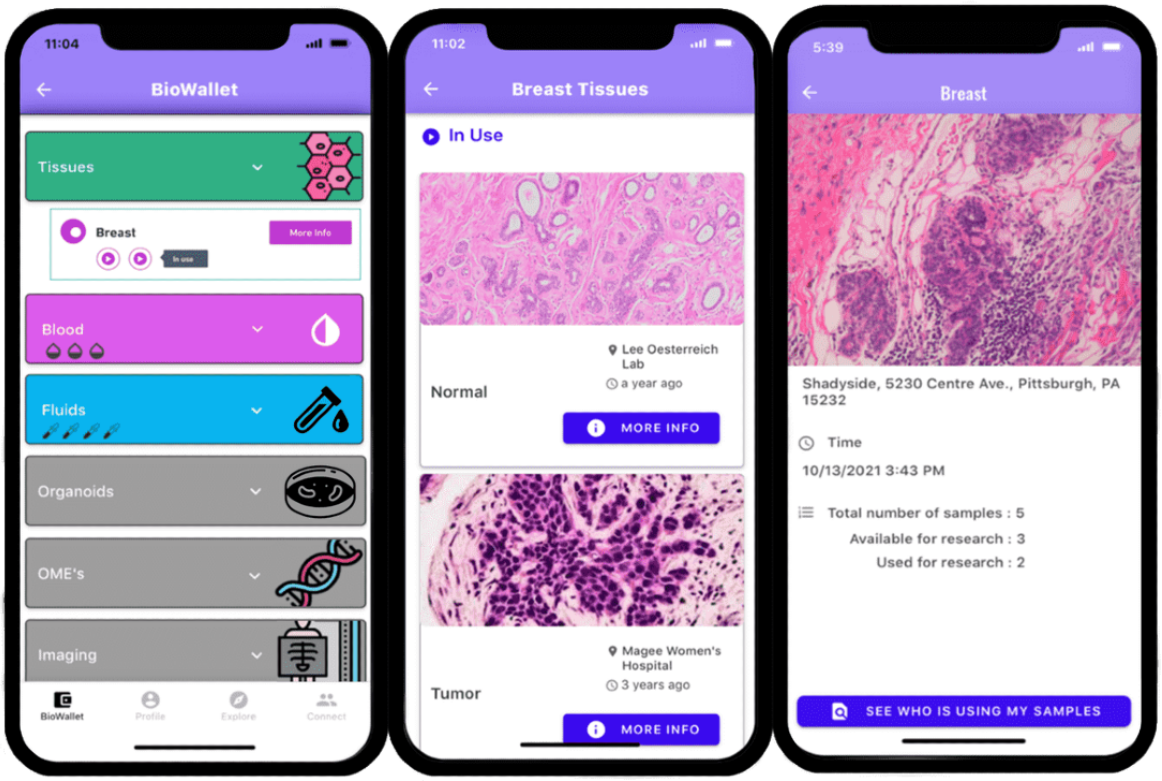
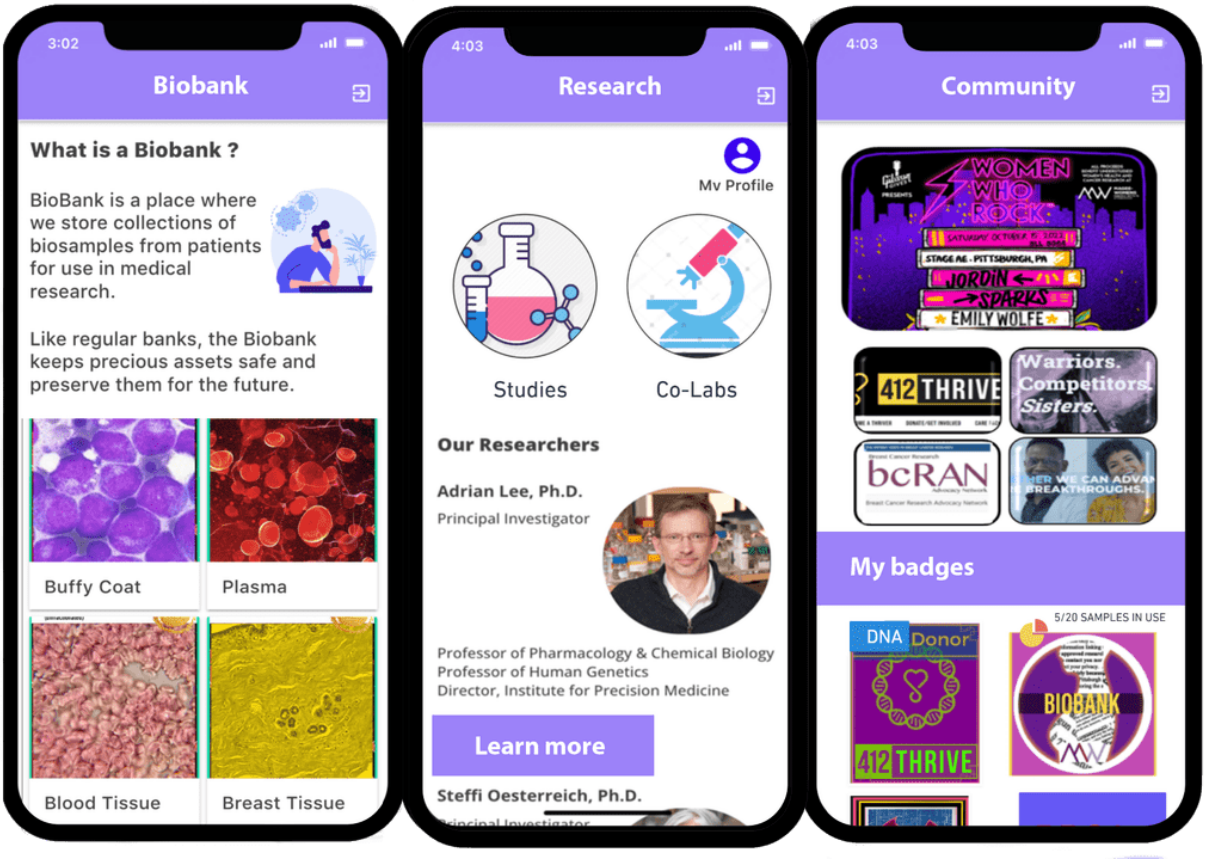


Figure 4. Biobank, research, and community screens. These screens, developed with Adalo, illustrate features explored in surveys, interviews, and workshops, including opportunities to learn about biobanking and specific specimens, follow scientists and learn about individual research programs, and connect with fellow donors and scientists.



Surveys and Semistructured Interviews

After developing the initial app screens, we embedded them into a survey to gather further evaluation and feedback from patients [29]. We conducted the survey in 2023 with a convenience sample, recruiting patients via flyers posted in breast care clinics across the health system. Participants took the survey on their own devices. IP addresses were used to identify duplicate entries from the same users. At the end of the survey, participants were invited to participate in an optional follow-up semistructured interview to discuss their survey answers. Survey content was iterated and rotated as app screens were finalized and in response to saturation of data from responses and follow-up interviews. Thus, survey data are presented with the denominator reflecting the number of participants who received each specific question and chose to reply. In total, 94 participants completed the survey, of whom 5 (5%) participated in the follow-up interviews (one participant was interviewed twice; once before and once after her cancer diagnosis). Demographics of the survey respondents, including age, race, and cancer stage, are presented in Table 1. The "Unknown" category represents participants who chose not to answer that question.

The survey asked participants their level of interest in the information presented on the app screens. After viewing the "biowallet" screens, participants were asked which of the presented data points were of interest to them. After viewing the "connect" screen, patients were asked whether identifying and belonging to a subcommunity of donors to the biobank would be of interest to them and, if so, how they might want to use the app to connect with others; for example, they were asked whether they were interested in connecting with researchers who used their samples from the biobank and, if so, what information they would like to exchange with them. They were also asked about their interest in connecting with other, similar patients who might be members of the same subcommunity (eg, have the same diagnosis code) and, if so, how.

In the follow-up semistructured interviews, participants and members of the research team viewed the participants' survey responses via screen sharing on an audio-recorded web-based call. The participant's survey responses were walked through one at a time, and the participant was asked to go into further detail about their reasoning for their answers. These interviews were meant to validate the survey questions by identifying any confusing language or images and to gain a deeper understanding of the motivations, preferences, interests, and concerns of participants. Each interview lasted 35 to 45 minutes. Interview participants were selected to represent key subsegments of the survey population, for example, age, education, and time from diagnosis.

Survey responses were analyzed, and summary statistics are presented. Transcripts of the follow-up interviews were coded for emergent themes in response to each set of screens. The results from both analyses informed changes to the presentation of biobank data and adjustments to the design of the screens to meet participant preferences and clarify content.

Table 1. Socioeconomic and clinical characteristics of survey respondents (n=94).

Demographics	Participants, n (%)[a]
Socioeconomic characteristics	
Age (y)	
<40	12 (13)
40-49	32 (34)
50-59	22 (23)
60-69	14 (15)
≥70	11 (12)
Unknown	3 (3)
Race	
White	64 (68)
Other (including multiracial)	10 (11)
Unknown	20 (21)
Household income (US $)	
<35,000	12 (13)
35,000-69,999	16 (17)
70,000-99,999	17 (18)
≥100,000	27 (29)
Prefer not to say	2 (2)
Unknown	20 (21)
Highest education	
Did not graduate high school	1 (1)
High school	6 (6)
Some college	9 (10)
Associate degree	10 (11)
Bachelor’s degree	21 (22)
Master’s degree	17 (18)
Doctoral or professional degree	9 (10)
Unknown	21 (22)
Clinical characteristics	
Breast cancer diagnosis	
Yes	56 (60)
No	35 (37)
Unknown	3 (3)
Cancer stage	
Unsure	5 (5)
0 or precancer	4 (4)
I	22 (23)
II or III	23 (24)
IV	2 (2)
Unknown	38 (40)
Time from diagnosis	
≥5 y ago	11 (12)

Demographics	Participants, n (%)[a]
>1 and <5 y ago	31 (33)
≤1 y ago	14 (15)
Unknown	38 (40)

[a]Due to rounding, not all percentages sum to 100%.

Design Workshop 1

We sought feedback on the "biowallet" screens, the "connect" screen, and 2 additional screens: "explore" and "profile." The "connect" feature was expanded to include 3 screens showing a list of all "lab" groups and subcommunities, a description and membership details of 1 specific "lab," and an action button to "add my samples" to a "lab." The screen called "explore" contained additional biobank-level data about the collection of biospecimens. The last addition, "profile," contained a hypothetical user-chosen avatar to represent themselves, as well as a collection of hypothetical achievement badges representing activities that users might be able to complete both in the app ("claimed my biowallet") and in research participation ("signed a consent form"). It also contained a progress bar indicating the users' completion of in-app tasks and achievements. We presented the second iteration of the screens to approximately 50 members of a local support group for patients with breast cancer in an embedded participatory design workshop. Members of the research team presented the screens, gave a short overview of their contents, and explained the overall intent of decentralized biobanking as well as the purpose of the dedicated app codesign workshop.

Participants provided feedback about the overall approach during a three-day participatory engagement event where the research team presented to and engaged with the local breast cancer patient and previvor population in an immersive setting. In this paper, we focus on describing results for 15 group representatives who attended a professionally facilitated in-person codesign workshop where they provided detailed feedback on individual app screen designs. Participants were given printed images of the screens, along with stickers—thumbs up or a heart to signify something they liked, a question mark to signify something they were confused by, and an exclamation point or a frowning face to signify something they did not like. Participants were asked to discuss the screens as a group and to place stickers on the printed images, either to signal their reaction to the entire screen or specific parts of the image. Some groups wrote on the images or in the margins to elaborate on their responses. As not all participants responded to all app screens, workshop data for all subsequently described workshops and cognitive walkthroughs are presented with the denominator reflecting the number of participants who provided written feedback on a given screen, page or design element.

The printouts were collected after the workshop, and the research team performed a content analysis of the sticker placements and notes left by participants. The identified themes were considered along with design research synthesis from expert consultants, as well as analysis of survey responses, semistructured interviews, and immersive engagement. Taken together, these themes informed further design changes that were then incorporated into a functioning mobile app prototype.

Phase 2

App Prototype Development

We developed a functional mobile app using Flutter [30]. The front-end patient-facing interfaces contained 4 sections: "biowallet," "explore," "connect," and "profile." We also developed a backend data architecture in which real biobank data could be imported into the app so that users could see personalized, real-time information about their biospecimens contained in the biobank collection. During the app's development, we consulted with institutional stakeholders to ensure regulatory compliance and compatibility with existing data resources to minimize burdens on researchers and data managers. Elsewhere, we describe a "digital honest broker" approach that enabled pilot participants to be connected directly to their own deidentified biospecimens [31].

Design Workshop 2

We held a second design workshop with a second local advocacy and support group for patients with any cancer type. A total of 25 participants attended. Members of the research team presented the app screens, gave an overview of their contents, and explained the intent of the workshop. Participants were then instructed to access the app on their personal devices and given a sample log-in leading to a deidentified demonstration account with biobank-level data and individual biospecimen data populated. Participants were given a sheet of paper with three columns—"Things I Liked," "Things I Didn't Like," and "Didn't Meet Expectations"—and were asked to write corresponding reactions as they explored the app freely by themselves or with one or two other participants. The papers were collected from nine small groups. Content analysis was performed to identify common areas of interest, enjoyment, displeasure, and dissatisfaction.

Cognitive Walkthroughs

During the second design workshop, members of the research team invited individuals at random to participate in a cognitive walkthrough to evaluate the usability of the application. Of the 25 workshop participants, 6 (24%) completed the walkthrough exercise. Participants were asked to perform several in-app tasks and seek specific information presented in the app while narrating out loud what they were doing. Members of the research team observed them using the app and took notes about their narration, evaluating their success in completing the assigned tasks.

We held seven additional cognitive walkthroughs with users recruited from participants in the phase 1 survey who agreed to be contacted about future research. These participants were

patients at the collaborating health system and were guided through the onboarding process that connected them to their own biospecimen information, if any existed. All users were able to see the "explore" and "profile" screens. Users who had also consented to the biobanking protocol (which was manually verified by the research team) were able to see the "connect" screen. If the user had samples in the biobank, they could see the "biowallet" screen populated with real data about their samples. Participants were led through the same cognitive walkthrough and asked to complete tasks and seek out specific information while they narrated their process as researchers observed and took notes.

Cognitive walkthroughs measure usability by collecting qualitative, experiential data from users directly. Researchers can infer usability by noting whether the users understand the task and are trying to achieve the right outcome, whether they take the appropriate action to achieve their intended outcome, and whether they correctly identify their success or the information sought from them. By having a mix of participants viewing uniform, hypothetical data and those viewing personalized data, we were able to gain generalized insights into usability, as well as insights into the personal value and added joy or frustration when engaging with one's own data. Both sets of participants were asked to create an account; perform the necessary authentications to populate their biospecimen data; identify how many of their samples had been used in research and how many were available; and to find a protocol that had used their samples, or, if their samples were unused, to find a protocol that studied their particular diagnosis.

Ethical Considerations

This study was approved by the University of Pittsburgh Institutional Review Board and quality improvement committees (IRB22010118, IRB22020035, and QRC3958). Participants provided informed consent before engaging with the study. The voluntary nature of the study was emphasized, including that their decision to participate in the study would have no impact on their care. Participants were able to skip questions or withdraw at any time. To safeguard privacy, data were stored on a secure institutional server, with survey responses deidentified and stored separately from personal information. Survey participants were paid US $10 for their participation upon completion. A subset of survey participants who agreed to participate in the follow-up interview were paid an additional US $20.

Results

Phase 1

Surveys

In reaction to the "biowallet" screen, participants were asked about their interest in using such an app to track their biospecimens donated to research. Of the 89 respondents, 20 (22%) were extremely interested, 22 (25%) were moderately interested, 24 (27%) were neutral or not sure, and 23 (26%) were not interested. Survey questions presented potential app features related to sample tracking, research feedback, and community engagement (Table 2), and follow up questions drilled down on preferences regarding specific design elements and user priorities.

We next assessed participants' interest in receiving updates regarding how, and by which researchers, their samples were used (Table 2). Participants (61/81, 75%) were most interested in learning "the outcomes of studies that use my tissues, like publications or new discoveries." Of these 61 participants, 14 (23%) provided qualitative details in optional open-ended follow-up questions, describing how or why they were interested in learning about research on their samples. For example, one respondent noted "I believe research participants should have full access to the data that is collected from them...The idea that my genetic data is setting in a lab for a variety of research studies, but I have no access to any of that data or the related information generated by that data is not fair or in my best interest."

The second highest area of interest concerned learning the status of their biospecimens regarding whether they had been used in research (45/81, 56%). Only 23% (19/81) of the participants were interested in seeing how many samples they had compared to other patients with the same diagnosis. Other data points of interest were the medical procedure from which their sample was collected (12/27, 44%), the current location of their sample (11/27, 41%), and collection details (when, volume, etc; 10/27, 37%).

Subsequently, we asked participants about their broader preferences for updates from researchers using their biospecimens with respect to scientific findings, new developments, and modes of communication (Table 2). Participants (47/60, 78%) expressed the greatest interest in receiving patient-friendly summaries of scientific information. Updates on new drugs or tests that become commercially available (39/60, 65%) and findings that may inform care (38/60, 63%) were also highly valued. Fewer patients were interested in updates on new publications (28/60, 47%), patient-friendly videos (22/60, 37%), study enrollment numbers (19/60, 32%), and new research grants (15/60, 25%).

In reaction to the "connect" screen, respondents expressed interest in connecting with other patients. Participants were most interested in connecting with patients with a similar type of breast cancer to theirs (41/83, 49%) and patients in their age group (36/83, 43%). Of the 83 respondents, 24 (29%) were not interested in connecting with any others. Table 2 details additional areas of interest.

Table 2. Interest in potential app features: biospecimen tracking, research updates, and community engagement[a].

App feature interest	Survey responses, n (%)
Sample tracking ("biowallet") features (n=81)	
Outcomes of the studies that use my tissues, like publications or new discoveries	61 (75)
Learning if/when my biospecimens are shared with researchers	45 (56)
Seeing how my own samples look under the microscope	35 (43)
Learning about the individual researchers who are studying my biospecimens	27 (33)
Seeing how many samples I have compared to patients with the same diagnosis as me	19 (23)
None	12 (15)
Other	3 (4)
Research updates ("scientist labs") (n=60)	
A patient-friendly summary of the scientific information	47 (78)
If there are new drugs or tests discovered that become commercially available products	39 (65)
If there are findings that might help my doctor and I make decisions about my care	38 (63)
When new publications come out (with access to that publication)	28 (47)
Short patient-friendly videos describing the research	24 (40)
How many patients are included	19 (32)
If there are new grants awarded to that research	15 (25)
Community engagement ("donor labs") (n=83)	
Similar type of breast cancer	41 (49)
Patients in my age group	36 (43)
Patients in the same research study	30 (36)
Similar stage of breast cancer	25 (30)
None	24 (29)
Similar genetic risk of cancer	18 (22)
Patient advocates	12 (14)

[a]These are not the only survey questions posed, but represent three key areas that were included in the survey. Where appropriate, questions were iterated to include screenshots of the app designs or concepts as they were refined.

Semistructured Interviews

While learning the outcomes of research on their biospecimens was identified as a high priority, concerns arose in the follow-up interviews. Of the 5 participants, 3 (60%) expressed concern about imposing additional burdens on busy scientists. They had expressed interest in learning about research on their specimens but clarified in interviews: "only if it doesn't take away from the work" and "I don't want to bother researchers, who are very busy" (Table 3). However, participants valued the opportunity to learn more about their samples; participant 3 expressed how specimen tracking can provide transparency about "what it's [biospecimen] being used for," implying how transparency may build trust.

In the follow-up interviews, 3 (60%) of the 5 participants expressed an interest in and need for communities that connect patients to research in general, stating that existing patient organizations were focused on emotional support, living with cancer, and survivorship rather than on biomedical research or the latest advances in treatments. Participant 2 expressed frustration with the lack of medical and scientific expertise in many patient support groups (Table 3): "...It's frustrating sometimes to get the kind of information you're looking for as a patient...with metastasis."

One participant undergoing diagnostic workup for cancer expressed interest in an additional support group, but without the granularity of connecting with other patients involved in the same studies. In contrast, one interview participant with a longer duration of cancer survivorship expressed that they did not have a need for additional patient connections. Multiple participants (3/5, 60%) expressed an interest in contributing additional information to scientists to provide additional context to their biospecimens. This included integrating data from genetic tests (eg, 23andMe), wearable devices (eg, Fitbit), and personal medical histories to support scientists. The motivation for doing so seemed to stem from the opportunity to "help...the scientific community, the research for cancer" (Participant 1) and elucidate "genetic links to breast cancer" (Participant 5). Participant 3 expressed interest in being informed about "a list of studies" they might qualify for, indicating a proactive approach to supporting biomedical research.

Table 3. Qualitative interview insights. The interview quotes are organized around feature categories of biobanking education, specimen tracking, connecting with others, and dynamic data sharing. Key standout phrases are italicized,

Proposed features ("naming convention") and interview quotes	Emerging themes
Learn biobanking ("biobank")	
• "I *totally geek out* on this stuff, so as much as you guys are willing to share and whatever level detail, I'm all about it." (Participant 1, Interview 2) • "I would want to know outcomes of the studies. I am kind of *a science geek*, like that's my interest, it's not the field that I'm in whatsoever, and I was not educated in that, but I'm very interested. I've read a lot about science and especially like bioscience, so I would say, I would also like to know outcomes, with the studies that use my tissues." (Participant 3)	Valuing scientific knowledge and approaches
• "Whatever you guys are willing to share is perfectly fine...*I just don't expect that detailed level of personal information* about the researchers, but if they're willing to share...it's fascinating...it's just not one of the absolutely necessary things, I feel." (Participant 1, Interview 2) • "*I don't expect to have the ability*, or even the need to identify specific researchers who I can go to. It's like as many people can use that tiny little tumor that you guys got out of me as possible it's fine, so what, for whatever purpose and whoever needs it, for whatever reason, is fine." (Participant 1, Interview 1) • "It was my pre-op...my surgeon, and a nurse, she just had a whole pile of papers. It [the biobanking consent form] *was just one more paper she needed to get signed*." (Participant 3)	Recognized unfamiliarity with biobanking, limited expectations and agency of donors within in current model
Specimen tracking ("biowallet")	
• "I don't need to know about all the researchers and everyone who's touching it. Just in general, like *what you're looking at and...what you're doing with it.*" (Participant 1, Interview 2) • "I don't feel the need to have control over any of that stuff—You guys can do with it what you see fit." (Participant 1, Interview 1) • "I *don't want to bother researchers*, who are very busy." (Participant 2) • "It would be interesting to see my samples, and my tissue, but not necessary, I guess. So, sort of like an added bonus." (Participant 5)	Balancing gaining personalized insights versus resource constraints for scientists
• "I mean, worst case scenario, like these are my few... last few minutes on earth... It would bring meaning...I'm *suffering but I'm suffering for a reason.*" (Participant 4) • "...People are fearful of having their information being used in nefarious ways; I think *this is a way of being super transparent and also involving the candidate*...like it's personal, it's something [very] personal, especially for women with breast tissue, very personal." (Participant 4)	Recognition for personalized contributions as a source of purpose and meaning
Community engagement ("labs")	
• "Probably in the lines of a support group kind of thing, it would be kind of neat to see...what other people were experiencing and going through, but not necessarily with regards to the study itself." (Participant 1, Interview 1) • "You have questions in there, 'Do you want to be connected with other donors?' and [in the survey] I said no, 'I don't use social media,' but I would use this, and perhaps if I do have to get surgery or I...perhaps [have cancer]... *it's like people who have dogs from the same litter, you know it's like, 'How's your pup doing?'*" (Participant 3) • "You know, now that I am talking this through, I would also...need...the type, the stage, the genetic risk of patients in my age group. I might want any one of those options. I am thinking futuristically now that it's become an actual reality or possibility...*I would be more open to connecting to others...beyond the same research.*" (Participant 3)	Some felt existing support groups were adequate, while others sought camaraderie and citizen science
• "One of the limitations with the patient advocacy groups that I've been in so far is that *there is limited expertise in the groups*...so it's frustrating sometimes to get information you're looking for as a patient...for example, drugs for preventing metastasis...what are the researchers recommending at this point [for people like me]." (Participant 2) • "I think the research is important and I'm just interested in knowing what research is going on, just because I think it's interesting, but *I wouldn't necessarily need to know what's happening with other patients* and their stuff."(Participant 5) • "I want to help, do whatever is possible to help research continue, but I don't feel the need to use something like this to connect with other people [i.e., patients]." (Participant 5)	Unmet need for direct connection with scientists, as opposed to other patients

Proposed features ("naming convention") and interview quotes	Emerging themes
Data sharing ("profile")	
• "I have the ancestry/23andMe, I do Fitbit stuff, so yeah, that'd be very cool to be able to incorporate all that...Anything that would help in any way... the scientific community, the research for cancer, anything, advanced in any way, for whatever reason, whatever purpose." (Participant 5) • "There's so much that's not known in terms of...genetic links to breast cancer...*I would share more information if it would help.*" (Participant 5) • "I would like to see a list of studies that I might qualify for, *I would provide details about my medical history or treatment.*" (Participant 3) • "[If] you guys need additional information or stats or blood or anything from me, I'm more than happy to help." (Participant 1, Interview 1)	Eager to contribute more specimens and data
• "I keep throwing out those ads for...'All of us' *because they won't give you anything...that's my reason for not participating*...[since] you have all my genetic information, you have no idea where it's going to be five years from now...But you won't let me have any access to it." (Participant 2) • "I think it's brilliant and I also think that *more people would participate if this were available*...Because I've had opportunities in the past, and...the biggest question that I have, and other people have, even when we're asked, 'Would you donate organs'...is the lack of knowledge about what it's being used for...that's what makes people suspicious and, especially, in an era where we both have like open source, you can find out information about everyone." (Participant 3)	Demand for greater reciprocity and accountability for data exchange

Design Workshop 1

Participants of the first design workshop (n=69) confirmed their interest in the concept of the app as well as the content of the app screens and provided valuable feedback on its presentation. A subset of 15 participants who participated in a dedicated codesign session documented their reactions to individual app screens. The "biobank" screens were appreciated by participants who liked the ability to view the entire collection of one's personal samples (7/15, 47%), as well as the sample-level details under the individual "biowallet" (4/15, 27%), with several emphasizing the linkage between samples and other medical or genetic data. Some were confused or upset by the technical details presented about each individual sample and about some of the terminology related to their biospecimens (eg, questioning what it means for a sample to be depicted as "in use"; 9/15, 60%). Of 15 co-designers, 4 (27%) noted that they would have preferred their samples to be organized chronologically, corresponding to their lived experience of appointments and procedures as specimen collection events, rather than initially stratified by sample type (tissue, blood, urine, organoids, etc.). For example, one co-designer emphasized: "Everything is timeline for us [patients with cancer]."

Moreover, 80% (12/15) of co-designers provided positive reactions to the "connect" screens, where they hearted the framing of virtual "lab" groups as 'patient-led' (5/12, 42%) and appreciated that there seemed to be actions that a user could take, such as joining a virtual "lab" and adding their samples to the group (6/12, 50%). Of the 12, 4 (33%) expressed confusion about the "lab" groups, including who is supposed to join them and for what purpose, while 3 (25%) expressed concerns about the legitimacy of *patient-led* "labs" because they might relate to actual future research on their samples, and they worried about the potential for unmoderated, off-topic, and illegitimate activity, with 1 participant stating "It should not feel like social media. To me it should feel more like science."

The "explore" screen, showing high-level information about the entire biobank collection of samples, prompted the most confusion. Of the 14 co-designers who were able to review this screen, 12 (86%) were confused by ≥1 terms on this page (eg, "consent" and "protocols"), and 2 (14%) suggested the addition of a glossary to help users understand. Other co-designers (3/14, 21%) incorrectly assumed that the "explore" page contained information about an individual's samples instead of the entire collection.

Participants also generally misunderstood the intended features of the "profile" page. Of the 11 co-designers who reviewed this page, 7 (64%) expressed confusion or dislike of the badges feature, with concerns that they were "useless," "not important," or excessively similar to features on social media sites, which they did not like. Finally, 18% (2/11) of the co-designers expressed concerns about the visibility of individual profile data, similarly worried about the parallels to existing social media, where all users can see individual profile pages.

Phase 2

App Prototype Development

Taking participant feedback from phase 1 activities, we developed a functional app prototype with 4 sections corresponding to the screen designs. The "biowallet" screens had been well received, and minimal changes in content were made. We removed the diagnosis code from the "biowallet" page to avoid triggering negative memories and emotions in users. We maintained sample organization by type (tissue, blood, urine, etc) to more closely resemble the structure of biobank data, where the type of sample and medium of preservation are more valuable than the collection date. However, to address participants' desire for a chronological arrangement, we listed

samples from the most recently collected to oldest *within* the sample type category. The "biowallet" page also showed a record of the users' informed consent to join the biobanking protocol. Users could click on individual samples to find information such as size, location, date collected, freezer number, and more. A generic picture of the corresponding sample type (eg, a picture of urine crystals under a microscope) accompanied each sample. Users could see whether each sample had been used in a research protocol or whether it was still available in the biobank.

The "connect" feature included patient-led "labs" organized by diagnosis, as in the original designs. We added a second class of "researcher-led labs" based on individual researchers who had used the patients' samples. Due to the structure of deidentification of the initial biobank data, the names of the researchers and details about their work were blinded, with just a numerical code representing individuals (eg, principal investigator 4); summary statistics of their work (eg, 6 protocols using 75 samples); and similar blinded, summary information about specific research protocols using patient samples (eg, "Your samples were used in protocol #4, which used 25 samples"). Importantly, displaying this information did not impose any additional workload or data input responsibilities on researchers. The patient-led and researcher-led "labs" featured some cross-referencing. A patient-led group with a shared diagnosis was linked to researcher "labs" studying this diagnosis and to protocols using samples with this diagnosis. Using existing databases, we sorted individual users into the appropriate patient-led and researcher-led "labs," adding a third section, "My Labs," where users could view just the groups of which they were members. These 2 new groups of "labs" were developed to satisfy phase 1 participants' enthusiasm for learning about research both in general and specifically about their samples, while maintaining maximum interoperability with existing biobank databases and protocols.

We updated terminology on the "explore" page in response to widespread confusion expressed by participants at the design workshop. Rather than relying on institutional terms such as "informed consent," we framed participation in the biobanking protocol as being a "member." For each sample category, we clarified how many patients were represented and how many samples there were in response to participants' confusion about whether they were looking at 1 person's samples on the "explore" page or the entire biobank's samples. The "explore" page contained expandable information about the entire biobank collection, organized by sample type (eg, breast tissue, blood, and urine).

On the "profile" page, users could view the deidentified code that links their account on the app to the biobank database, populating their sample information. Rather than showing a hypothetical set of badges that users might be incentivized to collect, which were not well received, we showed just 1 indicator of whether users had successfully been connected to their sample collection. On the "profile" page, we also developed short questionnaires about potential future research topics of interest. This feature served as proof of concept for how the app might be used to engage participants in research design or be used as a data collection tool.

Design Workshop 2

The second participatory design workshop aimed to further validate the clarity and value of the information presented in the app. In addition, we received feedback on the user experience of navigating the live app.

All 9 groups of patients participating in this workshop (consisting of 25 patients total) liked learning the information presented to them on the app. Of the 9 groups, 3 (33%) specifically mentioned the "real-time" nature of the information, with one-third (n=1, 33%) stating that they liked being able to "follow along" with research. Of the 9 groups, 5 (56%) noted that they would prefer or require more information before using the app, such as background materials or guided onboarding, to better understand the information presented; 7 (78%) wished for more information (eg, details about research protocols); and 4 (44%) wished for more capabilities (eg, a search bar or the ability to make their biospecimen data more or less visible to others).

Of the 9 groups, 7 (78%) noted that the app was easy to use. However, 7 (78%) of the 9 groups also noted usability and navigation as something they disliked. Of the 9 groups, 4 (44%) noted that the onboarding process—downloading the app, creating an account, logging in, and verifying the user's identity—did not meet their expectations or was not liked; 6 (67%) did not like a design feature of the app (eg, the choice of icon and the order in which the screens appeared); and 3 (33%) listed unmet accessibility expectations, including the desire for use on a larger screen or the preference for a web platform rather than a mobile app.

Echoing concerns from the first workshop, of the 9 groups, 3 (33%) were dissatisfied with the lack of explicit security assurances or data use terms, while 1 (11%) liked the fact that users' identities were not featured on the app and were instead represented by a unique deidentified code.

Cognitive Walkthroughs

Cognitive walkthroughs with 13 patients also gauged the usability of the app. The walkthroughs illuminated an onerous and confusing onboarding process, where participants had to create an account and submit information to link their biobank data, after which they received a notification and request for more information once their biospecimen data were populated in the app. For users who were participants in the biobank, the interim waiting period for biospecimen data was as long as 1 week, and users often mistakenly thought they had done something incorrectly. Nearly all walkthrough participants expressed confusion and uncertainty about the onboarding and biospecimen data verification process, even when completed successfully.

Once they successfully logged in and were able to view biospecimen data, most of the participants successfully described their biospecimen data, including how many of their samples were available for research. Those participants who were members of the biobank, viewing their own biospecimen data, were curious as to why their samples had not been used and expressed a willingness to take action to ensure that their samples were used in research (eg, by offering to donate more

samples or to pool samples with other participants). Overall, the success rates of completing this task indicated the usability of the app. Lower usability, namely the inability to identify successful completion of the task, arose in specific contexts applicable only to those participants who were attempting to view their own biobank data. Of the 13 participants, 3 (23%) had consented to participate in biobanking, but none of their clinical biospecimens made it to the biobank. Thus, their "biowallet" page showed no samples, although their "membership" was verified. This was a source of confusion, and participants incorrectly assumed that they had made a mistake in the onboarding process.

Participants were able to correctly identify a protocol that had used their samples or that had studied their disease. These participants successfully navigated from the "biowallet" screen to the "connect" screen and correctly understood the information presented about the protocol they identified. Some of the participants (7/13, 54%) expressed disappointment in the limited amount of information presented about the protocol or the associated researcher. Consistent with the initial survey findings, participants were most interested in learning the details and outcomes of the research conducted on their samples or their disease. In sum, the cognitive walkthroughs showed the usability of and interest in the app and revealed the need to improve the onboarding process and to better explain less intuitive scenarios, such as having no banked samples as a consented biobanking participant. The walkthroughs also confirmed participant interest in learning what happened to their biospecimens and about research on banked biospecimens.

Discussion

Key Findings

Throughout the design process, participants indicated a strong interest in and demand for learning about biobanking. They found intuitive value in the ability to track their own biospecimens. Indeed, participants positively received almost all of the information presented and commonly expressed a desire for more. In all phases, participants were most interested in learning about the research being conducted on their disease or specifically using their biospecimens, including the results of this research. They noted that they were not able to find this information on their own and were not able to satisfy general curiosity about research by participating in existing patient advocacy and support groups. These findings add to existing literature confirming public interest in greater transparency and engagement in biomedical research [32-36].

Mixed responses about the usability of the app demand further research and attention. Overall, participants in all stages were interested in using the app, although the content and presentation of information needed improvement. Some of the participants sought language and design similar to that of their medical records. Presenting research information in this way may make it more intuitively understandable but may also contribute to therapeutic misconception by misleading participants into believing that the research information and participation are of a similar clinically actionable quality. Thus, careful onboarding and framing of research information must ensure participant understanding and offer opportunities to overcome confusion.

Participants expressed a demand for a "legitimate" platform that did not suffer from what they saw as the risks and ills of social media: data misuse, a lack of privacy, and the potential for negativity or misinformation. While mobile phones and apps are widely used [37], concerns about mobile apps may make them an unacceptable platform for some. An additional facet of building trust with participants can be achieved through transparency surrounding how their biospecimens are being used. An interview participant expressed the following:

> *The biggest question that I have and other people have even when we're asked you know, would you donate organs...is...the lack of knowledge about what it's being used for.*

This approach to rebuilding trust reflects ongoing movements in biomedical research aimed at improving transparency and accountability, such as interactive and accessible "data walks" designed to engage communities and stimulate dialogue [38].

However, the de-identification of biospecimens presents challenges for direct transparency and engagement. Other challenges exist as well—meaningful transparency and engagement with patients is time and labor intensive [39,40]. Even when community engagement is required or encouraged, there are few tools or guidelines for success, and interactions may be transient and transactional [40]. Biobanking and scientific research are deeply technical disciplines that historically have minimally interfaced with the public, and require extensive time, skill, and technology to translate into accessible, meaningful plain language [39]. Finding relevant patient communities and sustaining meaningful, collaborative relationships with them might prove beneficial to biobanks, or scientists themselves, but this is not required, supported, or incentivized by traditional academic reward systems or many institutions [41-43]. Engaging patients in biobanking might fulfill obligations of respect; however, evidence of its long-term impact on research participation, understanding, and attitudes is the subject of ongoing work [40].

Our experience confirms the need for specialized skills and dedicated time and resources to deliver transparency and engagement effectively and safely. New technologies, eg blockchain and generative artificial intelligence, are essential for these endeavors to be more efficient and scalable, but their development remains time and labor intensive. Participants were largely unfamiliar with biobanking, research processes or related jargon. Thoughtful, readily digestible and personalized education about each unique research context may be a critical aspect of the app onboarding process. Participants were confused about less intuitive but common scenarios, such as having consented to biobanking but having no samples in the bank. While this confusion may stem from a lack of recall or understanding of the initial informed consent for biobanking, delivering follow-up information about actual participation (or lack thereof) presents new risks of misunderstanding that must be managed.

While our study focused on patients with breast cancer, the demand for transparency and engagement may extend to other medical fields. We found that the survey participants (61/68, 90%) were most interested in learning the outcomes of research using their tissue samples. Thus, conditions with active research and strong patient interest in new treatment paradigms, such as additional cancer types and sickle cell disease, are potential areas where this platform may be effective. Further research is needed to assess the feasibility, acceptability, and effectiveness of this platform across diverse medical contexts and patient demographics.

Finally, in light of the strong reception and demand by participants, it must be acknowledged that returning deidentified, individual-level research information to participants is uncommon and unaccommodated by existing paradigms. Existing data architectures are built for a research enterprise from which participants are intentionally excluded. Working to disrupt the paradigm by allowing participants to view research data and activities presented both technological and regulatory implications. While participants were unaccustomed to receiving this type of information and, as a result, sometimes struggled to understand it, biobanks were also unaccustomed to sharing this type of information and, as a result, were sometimes hesitant to do so. Any technological tool aiming to enable transparency and engagement in research must be acceptable to all stakeholders. Careful and inclusive co-design is imperative.

Limitations and Next Steps

The patient population we engaged in this work consisted of only women due to the nature of breast disease. Our participants were mostly White (64/94, 68%) and were more educated and had higher incomes than the national average, although they more closely resembled local demographics. Our participant population may also reflect self-selection bias—those interested in learning about research on their biospecimens or in general may have been more likely to participate in any research, including ours. Our technology was intended for piloting among this group and was therefore tailored to their preferences. For these reasons, further research will be necessary to validate interest in and usability of the technology in other patient populations (eg, men and healthy individuals).

Moreover, most of the patients we engaged with throughout all stages of design and development were unfamiliar with the term "biobank." This underscores a key challenge associated with delivering an accessible and patient-friendly biobanking app. This variation in patient knowledge regarding biobanking, even among those who already consented to biospecimen donation, limits the ability to fully capture and address diverse patient preferences.

Across the surveys and workshops, questions and app content were iterated upon as app screens were finalized, new topics were raised by participants, and data saturation was reached through survey responses and follow-up interviews. As a result, the sample size of participants exposed to specific materials varied, leading to disproportionately lower sample sizes for certain areas of inquiry. Additional work is needed to investigate a larger sampling of patients' design preferences. New technologies can facilitate meaningful transparency in biomedical research to patients in a personalized yet efficient way. This research showed that patients desire transparency and revealed the specific information they were interested in learning about their biospecimens. By deeply engaging patients and other stakeholders in participatory methods, siloed and specialized research data can be translated into meaningful patient engagement. We worked to design a mobile app that allows patients with breast cancer to track and learn about their donated biospecimens. Surveys, interviews, design workshops, and cognitive walkthroughs helped to make the app maximally valuable and usable for patients while maintaining efficient integration with existing data assets.

By creating this decentralized biobanking app, we built a platform upon which methods of transparency and engagement can be tested in subsequent pilots; for example, further research can measure the app's impact on patients' willingness to consent to biobank research, to remain enrolled and engaged longitudinally, and to actively participate in community-engaged research.

Acknowledgments

The authors thank Diana Mendoza Cervantes, Sneha Vaidhyam, and Dmitriy Babichenko, as well as Raelynn O'Leary and Ashley Deal of Dezudio, for their contributions to study design and execution. The authors also thank Sam Moatz, Jen Kehm, the members of 412Thrive, and the members of Cancer Bridges. Finally, the authors express utmost gratitude to the participants of this study: "Your openness, honesty, and trust motivate all that we do."

Conflicts of Interest

AH, WS, and MG are consultants for de-bi, co. All other authors declare no conflicts of interest.

References

1. Boyer GJ, Whipple W, Cadigan RJ, Henderson GE. Biobanks in the United States: how to identify an undefined and rapidly evolving population. Biopreserv Biobank 2012 Dec;10(6):511-517 [FREE Full text] [doi: 10.1089/bio.2012.0034] [Medline: 24845137]
2. Henderson GE, Cadigan RJ, Edwards TP, Conlon I, Nelson AG, Evans JP, et al. Characterizing biobank organizations in the U.S.: results from a national survey. Genome Med 2013 Jan 25;5(1):3 [FREE Full text] [doi: 10.1186/gm407] [Medline: 23351549]

3. Tarling TE, Byrne JA, Watson PH. The availability of human biospecimens to support biomarker research. Biomark Insights 2022 Apr 19;17:11772719221091750 [FREE Full text] [doi: 10.1177/11772719221091750] [Medline: 35464611]
4. Li Y, Tang P, Cai S, Peng J, Hua G. Organoid based personalized medicine: from bench to bedside. Cell Regen 2020 Nov 02;9(1):21 [FREE Full text] [doi: 10.1186/s13619-020-00059-z] [Medline: 33135109]
5. Barnes RO, Watson PH. Precision medicine: driving the evolution of biobanking quality. Healthc Manage Forum 2020 May;33(3):102-106. [doi: 10.1177/0840470419898874] [Medline: 31957497]
6. Olson JE, Bielinski SJ, Ryu E, Winkler EM, Takahashi PY, Pathak J, et al. Biobanks and personalized medicine. Clin Genet 2014 Jul;86(1):50-55 [FREE Full text] [doi: 10.1111/cge.12370] [Medline: 24588254]
7. Zhou Z, Cong L, Cong X. Patient-derived organoids in precision medicine: drug screening, organoid-on-a-chip and living organoid biobank. Front Oncol 2021;11:762184 [FREE Full text] [doi: 10.3389/fonc.2021.762184] [Medline: 35036354]
8. Castillo-Pelayo T, Babinszky S, LeBlanc J, Watson PH. The importance of biobanking in cancer research. Biopreserv Biobank 2015 Jun;13(3):172-177. [doi: 10.1089/bio.2014.0061] [Medline: 26035006]
9. Warner TD, Weil CJ, Andry C, Degenholtz HB, Parker L, Carithers LJ, et al. Broad consent for research on biospecimens: the views of actual donors at four U.S. medical centers. J Empir Res Hum Res Ethics 2018 Apr;13(2):115-124 [FREE Full text] [doi: 10.1177/1556264617751204] [Medline: 29390947]
10. Sobel ME, Dreyfus JC, Dillehay McKillip K, Kolarcik C, Muller WA, Scott MJ, et al. Return of individual research results: a guide for biomedical researchers utilizing human biospecimens. Am J Pathol 2020 May;190(5):918-933 [FREE Full text] [doi: 10.1016/j.ajpath.2020.01.014] [Medline: 32201265]
11. Tomlinson T, De Vries RG. Human biospecimens come from people. Ethics Hum Res 2019 Mar;41(2):22-28 [FREE Full text] [doi: 10.1002/eahr.500009] [Medline: 30895751]
12. Rothstein MA. Is deidentification sufficient to protect health privacy in research? Am J Bioeth 2010 Sep;10(9):3-11 [FREE Full text] [doi: 10.1080/15265161.2010.494215] [Medline: 20818545]
13. Gross MS, Hood AJ, Rubin JC, Miller RCJ. Respect, justice and learning are limited when patients are deidentified data subjects. Learn Health Syst 2022 Jul 04;6(3):e10303 [FREE Full text] [doi: 10.1002/lrh2.10303] [Medline: 35860318]
14. Mamo N, Martin GM, Desira M, Ellul B, Ebejer JP. Dwarna: a blockchain solution for dynamic consent in biobanking. Eur J Hum Genet 2020 May;28(5):609-626 [FREE Full text] [doi: 10.1038/s41431-019-0560-9] [Medline: 31844175]
15. Gross MS, Miller RC. Ethical implementation of the learning healthcare system with blockchain technology. Blockchain Healthc Today 2019 Jun 07;2. [doi: 10.30953/bhty.v2.113]
16. Aguayo GA, Goetzinger C, Scibilia R, Fischer A, Seuring T, Tran VT, et al. Methods to generate innovative research ideas and improve patient and public involvement in modern epidemiological research: review, patient viewpoint, and guidelines for implementation of a digital cohort study. J Med Internet Res 2021 Dec 23;23(12):e25743 [FREE Full text] [doi: 10.2196/25743] [Medline: 34941554]
17. The James Lind Alliance guidebook. James Lind Alliance. 2021 Mar. URL: https://www.jla.nihr.ac.uk/jla-guidebook [accessed 2025-02-04]
18. Barony Sanchez RH, Bergeron-Drolet LA, Sasseville M, Gagnon MP. Engaging patients and citizens in digital health technology development through the virtual space. Front Med Technol 2022 Nov 25;4:958571 [FREE Full text] [doi: 10.3389/fmedt.2022.958571] [Medline: 36506474]
19. Antoine-LaVigne D, Hayes T, Fortenberry M, Ohikhuai E, Addison C, Mozee SJ, et al. Trust and biomedical research engagement of minority and under-represented communities in Mississippi, USA. Int J Environ Res Public Health 2023 Jan 16;20(2):1592 [FREE Full text] [doi: 10.3390/ijerph20021592] [Medline: 36674346]
20. Scharff DP, Mathews KJ, Jackson P, Hoffsuemmer J, Martin E, Edwards D. More than Tuskegee: understanding mistrust about research participation. J Health Care Poor Underserved 2010 Aug;21(3):879-897 [FREE Full text] [doi: 10.1353/hpu.0.0323] [Medline: 20693733]
21. Kass NE, Faden RR. Ethics and learning health care: the essential roles of engagement, transparency, and accountability. Learn Health Syst 2018 Oct 18;2(4):e10066 [FREE Full text] [doi: 10.1002/lrh2.10066] [Medline: 31245590]
22. Hood A, Macis M, Cervantes DM, Bear T, Khan J, Atkinson JM, et al. Privacy, policy, and profits: survey of patient values and preferences for research on de-identified biosamples. Advance Preprint posted online on January 30, 2024. [doi: 10.31124/advance.24155085.v1]
23. Gross M, Miller RC. Protecting privacy and promoting learning: blockchain and privacy preserving technology should inform new ethical guidelines for health data. Health Technol 2021 Aug 20;11(5):1165-1169. [doi: 10.1007/s12553-021-00589-9]
24. Gross MS, Hood AJ, Miller RCJ. Nonfungible tokens as a blockchain solution to ethical challenges for the secondary use of biospecimens: viewpoint. JMIR Bioinform Biotechnol 2021 Oct 22;2(1):e29905 [FREE Full text] [doi: 10.2196/29905] [Medline: 38943235]
25. Spinuzzi C. The methodology of participatory design. Tech Commun 2005 May;52(2):163-174 [FREE Full text] [doi: 10.4135/9781529696202.n4]
26. Cavagnuolo M, Capozza V, Matrella A. The walkthrough method: state of the art, innovative aspects, and application fields. In: Handbook of Research on Advanced Research Methodologies for a Digital Society. Hershey, PA: IGI Global Scientific Publishing; 2022.

27. Whimsical homepage. Whimsical. URL: https://whimsical.com/ [accessed 2025-02-21]
28. Adalo homepage. Adalo. URL: https://www.adalo.com/ [accessed 2025-02-21]
29. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
30. Flutter homepage. Flutter. URL: https://flutter.io/ [accessed 2025-02-21]
31. Sanchez W, Linder L, Miller RC, Hood A, Gross MS. Non-fungible tokens for organoids: decentralized biobanking to empower patients in biospecimen research. Blockchain Healthc Today 2024;7 [FREE Full text] [doi: 10.30953/bhty.v7.303] [Medline: 38715762]
32. Jones RD, Krenz C, Griffith KA, Spence R, Bradbury AR, De Vries R, et al. Governance of a learning health care system for oncology: patient recommendations. JCO Oncol Pract 2021 Apr;17(4):e479-e489. [doi: 10.1200/op.20.00454]
33. Peppercorn J, Campbell E, Isakoff S, Horick NK, Rabin J, Quain K, et al. Patient preferences for use of archived biospecimens from oncology trials when adequacy of informed consent is unclear. Oncologist 2020 Jan;25(1):78-86. [doi: 10.1634/theoncologist.2019-0365] [Medline: 31492767]
34. Sanderson SC, Brothers KB, Mercaldo ND, Clayton EW, Antommaria AH, Aufox SA, et al. Public attitudes toward consent and data sharing in biobank research: a large multi-site experimental survey in the US. Am J Hum Genet 2017 Mar 02;100(3):414-427 [FREE Full text] [doi: 10.1016/j.ajhg.2017.01.021] [Medline: 28190457]
35. Kaufman DJ, Murphy-Bollinger J, Scott J, Hudson KL. Public opinion about the importance of privacy in biobank research. Am J Hum Genet 2009 Nov;85(5):643-654 [FREE Full text] [doi: 10.1016/j.ajhg.2009.10.002] [Medline: 19878915]
36. D'Abramo F, Schildmann J, Vollmann J. Research participants' perceptions and views on consent for biobank research: a review of empirical data and ethical analysis. BMC Med Ethics 2015 Sep 09;16:60 [FREE Full text] [doi: 10.1186/s12910-015-0053-5] [Medline: 26354520]
37. State of mobile 2024. Sensor Tower. URL: https://sensortower.com/state-of-mobile-2024 [accessed 2025-02-04]
38. Fattah L, Johnson J, Clark U, Hess L, Palermo AG, Figueroa Acosta DM, et al. Building trust and transparency in biomedical sciences through data walks. Nat Med 2024 Aug 31;30(8):2103-2104. [doi: 10.1038/s41591-024-03042-2] [Medline: 38822130]
39. Coors ME, Westfall N, Zittleman L, Taylor M, Westfall JM. Translating biobank science into patient-centered language. Biopreserv Biobank 2018 Feb;16(1):59-63 [FREE Full text] [doi: 10.1089/bio.2017.0089] [Medline: 29190122]
40. Luger TM, Hamilton AB, True G. Measuring community-engaged research contexts, processes, and outcomes: a mapping review. Milbank Q 2020 Jun;98(2):493-553 [FREE Full text] [doi: 10.1111/1468-0009.12458] [Medline: 32428339]
41. Stires H, Bado I, Brown T, Carlson M, Chan IS, Echeverria GV, et al. Improving the odds together: a framework for breast cancer research scientists to include patient advocates in their research. NPJ Breast Cancer 2022 Jun 30;8(1):75 [FREE Full text] [doi: 10.1038/s41523-022-00440-y] [Medline: 35773258]
42. Klingler C, von Jagwitz-Biegnitz M, Baber R, Becker KF, Dahl E, Eibner C, et al. Stakeholder engagement to ensure the sustainability of biobanks: a survey of potential users of biobank services. Eur J Hum Genet 2022 Dec;30(12):1344-1354 [FREE Full text] [doi: 10.1038/s41431-021-00905-x] [Medline: 34031552]
43. Spector-Bagdady K, De Vries RG, Gornick MG, Shuman AG, Kardia S, Platt J. Encouraging participation and transparency in biobank research. Health Aff (Millwood) 2018 Aug;37(8):1313-1320 [FREE Full text] [doi: 10.1377/hlthaff.2018.0159] [Medline: 30080467]

Edited by A Kushniruk, E Borycki; submitted 12.04.24; peer-reviewed by R Rahimi, M Bagewadi Ellur, L Weinert; comments to author 13.09.24; revised version received 23.12.24; accepted 14.01.25; published 05.03.25.

Please cite as:
Dewan A, Eifler M, Hood A, Sanchez W, Gross M
Building a Decentralized Biobanking App for Research Transparency and Patient Engagement: Participatory Design Study
JMIR Hum Factors 2025;12:e59485
URL: https://humanfactors.jmir.org/2025/1/e59485
doi:10.2196/59485
PMID:40053747

Original Paper

Impact of Bottom-Up Cocreation of Nursing Technological Innovations: Explorative Interview Study Among Hospital Nurses and Managers

Saskia van Steenis[1], MSc; Onno Helder[2,3], PhD; Helianthe S M Kort[4,5], PhD; Thijs van Houwelingen[4], PhD

[1]Nursing Sciences, program in Clinical Health Sciences, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

[2]Department of Create4Care, Erasmus MC, University Medical Center Rotterdam, Rotterdam, The Netherlands

[3]Research Centre Innovations in Care, Rotterdam University of Applied Sciences, Rotterdam, The Netherlands

[4]Research Group Technology for Healthcare Innovations, Research Centre for Healthy and Sustainable Living, University of Applied Sciences Utrecht, Utrecht, The Netherlands

[5]Building Healthy Environments for Future Users Group, Department of the Built Environment, Eindhoven University of Technology, Eindhoven, The Netherlands

Corresponding Author:
Thijs van Houwelingen, PhD
Research Group Technology for Healthcare Innovations, Research Centre for Healthy and Sustainable Living
University of Applied Sciences Utrecht
Heidelberglaan 7
Utrecht, 3584 CS
The Netherlands
Phone: 31 641097762
Email: thijs.vanhouwelingen@hu.nl

Abstract

Background: In health care, the use of nursing technological innovations, particularly technological products, is rapidly increasing; however, these innovations do not always align with nursing practice. An explanation for this issue could be that nursing technological innovations are developed and implemented with a top-down approach, which could subsequently limit the positive impact on practice. Cocreation with stakeholders such as nurses can help address this issue. Nowadays, health care centers increasingly encourage stakeholder participation, which is known as a bottom-up cocreation approach. However, little is known about the experience of nurses and their managers with this approach and the innovations it results in within the field of nursing care.

Objective: This study aims to explore nurses' and their managers' experiences with a bottom-up cocreation approach in order to assess the impact of this way of working and the resulting nursing technological innovations in an academic hospital. This insight can also inform decisions on whether the bottom-up cocreation approach should be more widely disseminated.

Methods: A qualitative study using semistructured interviews was conducted with 15 participants, including cocreator nurses, end-user nurses, and their managers. First, the data were thematically analyzed. In addition, a strengths, weaknesses, opportunities, and threats analysis was conducted.

Results: The various experiences of the participants were described in 3 main themes: enhanced attractiveness of the nursing profession, feeling involved due to a cocreation environment, and experienced benefits and challenges in using cocreated products. In addition, numerous strengths and opportunities perceived by the participants were identified as associated with the bottom-up cocreation approach and resulting useful products within nursing care; for example, cocreation contributed to job satisfaction and substantially contributed to the ease of use of the innovations that were developed.

Conclusions: The findings underscore that cocreation with nurses enhances the appeal of the nursing profession and aligns nursing technological innovations with practical nursing challenges. Embracing a culture of cocreation has the potential to foster a culture of continuous improvement and innovation in nursing care.

(JMIR Hum Factors 2025;12:e60543) doi:10.2196/60543

KEYWORDS

stakeholder participation; cocreation; nursing; innovation; bottom-up approach; diffusion of innovation; qualitative research; nurses

Introduction

Background

Nowadays, technological products and other innovations are being used more frequently in the health care sector [1-3] to minimize physical strain, such as lifting patients or transporting beds, and encourage ergonomic approaches during nursing tasks. Nurses need to use these innovations during patient care to manage various health care challenges, such as the increasing demand and complexity of health care [4,5]. An example of a technological product, which will be referred to as a "nursing technological innovation" in this paper, is the TrulyEasy clamp, designed for convenient 1-handed adjustment of the arterial line to the correct height [3]. Another example of a technological product is LineConnector, a product designed to efficiently organize infusion lines and prevent tangling [3].

Unfortunately, technological solutions aimed at solving nursing issues do not always align with daily nursing practice [2,6]. A potential explanation for this lies in the top-down approach used during the development and implementation of many innovations. Policy makers may prioritize regulatory and payment interests, while technicians focus on the functionality of innovations, without fully understanding the user perspectives and the nursing context [6-8]. Consequently, some solutions may inadequately align with the needs of the nursing profession, leading to limited impact on nursing practice [2,6]. Improving the adoption of innovations in nursing practice can be addressed in the development and implementation phase of these innovations through stakeholder participation [1,5,9-12].

A growing number of health care centers are facilitating an environment that allows stakeholders to participate in the development and implementation of nursing technological innovations [1,6,9,10,12]. This involvement, also called cocreation [13], with a bottom-up approach [8], involves an interaction in which various stakeholders work together to produce an outcome that is valued by all stakeholders involved [13]. The bottom-up approach allows stakeholders to share their knowledge or opinions about issues presented in daily activities and suggest potential solutions [8,14]. This approach can support the development of nursing technological innovations that may have an impact on nursing care [1,5,9-12]. FlushEgg is an example of a cocreated nursing technological innovation. Nurses reported experiencing challenges and discomfort while flushing the narrow central lines in neonates. Therefore, a syringe attachment was designed to enhance comfort in the palm during the application of pressure [3]. Another example of a cocreated nursing technological innovation is KoosGuard, which includes a holder that can be placed in the bed through which the cables and infusion lines are guided without becoming tangled or pinched [3].

However, the bottom-up approach requires nurses to invest time, which is challenging considering many nurses face time constraints [15,16]. If these nursing technological innovations lack the required quality, this could result in decreased application in nursing practice [2,6]. Furthermore, a poorly developed nursing technological innovation could lead to resistance from nurses, resulting in a negative attitude toward these innovations [9,17]. To fully use the potential of innovations in nursing practice, the innovations need to be adequately designed, which requires the involvement of nurses [9,18]. Despite the increasing involvement of nurses in the cocreation of nursing technological innovations, little is known about how this approach and its resulting innovations are experienced by nurses. Therefore, this study seeks to explore the experience resulting from the bottom-up approach of cocreation with nurses and the resulting nursing technological innovations developed through this approach. The perspectives of nurses, both as cocreators and end users, and their managers are key to understanding the impact of this collaborative process. These results could offer a deeper insight into how nurses, both as cocreators and end users, along with their managers, perceive the bottom-up cocreation approach and its resulting nursing technological innovations. A comprehensive understanding of their experiences, both positive and negative, can significantly contribute to the overarching goal of enhancing patient care and nursing practice, making it safer and more effective.

Aim

This study explores nurses' and their managers' experiences with a bottom-up cocreation approach in order to assess the impact of this way of working and the resulting nursing technological innovations in an academic hospital. This insight can also inform decisions on whether the bottom-up cocreation approach should be more widely disseminated.

Methods

Study Design

This study used an exploratory, descriptive, qualitative research design using individual interviews conducted at an academic hospital in the Netherlands between February and April 2023. This design facilitated the exploration of participants' experiences, which were further analyzed through a thematic analysis and subsequently through a strengths, weaknesses, opportunities, and threats (SWOT) analysis [19]. The methods and subsequent results were reported according to the Consolidated Criteria for Reporting Qualitative Research (COREQ) checklist [20].

Population and Setting

The study population consisted of three subgroups: (1) nurses with experience in cocreation, (2) nurses with experience as end users, and (3) health care managers from the nursing departments at an academic hospital where the innovations were implemented. In addition, all eligible nurses and managers were required to have at least 1 year of experience within the nursing department where the innovations were implemented. The 1-year experience criterion ensured that nurses and managers were

sufficiently familiar with the innovations and context. The setting comprised an innovation department established within the academic hospital.

Sampling Strategy and Recruitment

To gain more understanding and in-depth information from a broader perspective, a heterogeneous group of nurses and their health care managers were included through purposive sampling [19,21,22]. Potential nurse cocreators and managers for this study were recruited and informed via an email by OH about participating in the interviews. Interested individuals were required to actively contact SvS by email. Upon receiving an email, SvS responded with a participant information letter and an informed consent form. Once participants agreed to participate in the study, interviews were scheduled at a mutually convenient time. For the nurse end users, SvS visited the nursing departments where nursing technological innovations were integrated into nursing practice. Interested nurse end users were able to participate in an interview during their shifts at a time convenient for them. The nurse end users were informed via a participant information letter. After informed consent was given, the interviews started. The face-to-face interviews were conducted at the academic hospital.

Sample Size

A total of 14 to 20 participants were required to obtain saturation of knowledge with a heterogeneous sample group [19]. Therefore, the aim of this study was to include at least 14 participants [19]. The study involved 15 participants, including 10 (67%) women and 5 (33%) men.

Data Collection

Data were collected through semistructured interviews using an interview guide based on literature and expert opinions [19]. The interview guide covered various topics (Table 1), including the participants' attitude [23] toward technology, the bottom-up cocreation approach, and the nursing technological innovations developed through this approach. It also explored the acceptance [9,10,23], usability [12,23], and adoption [10,23] of these innovations in nursing care as well as their impact on patient care and nursing practice [1,11,12,24]. Demographic characteristics (age, gender, and specialization), along with the adoption rate, provided an overview and background information of the participants in this study. The adoption rate proposed by Rogers [25] allowed the measurement of individuals' attitudes toward innovations, providing valuable insights into how the participants generally perceive and approach innovations. The adoption rate according to Rogers [25] was measured by asking the participants which category they identified with. They could choose one of the following five categories: (1) driven by change and introduction of innovations, (2) leading in adopting innovations, (3) deliberately adopting innovations, (4) adoption is an economic necessity and a response to social peer pressure, and (5) have a traditional view and are more skeptical about innovations [25]. A pilot test was conducted to refine the interview guide and involved expert assessment and field testing with one participant. After the interviews, the first author created memos recording thoughts that arose due to the interviews [19]. The interviews were audio recorded and conducted by SvS, a nurse with no prior experience at the hospital where the research took place. SvS was trained in conducting interviews. Throughout the data collection process, discussions were held with the research team, including investigators SvS, TvH, and OH, to refine the interview questions and ensure that they effectively addressed the research question.

Table 1. Interview guide topics and prompt questions.

Topics	Prompt questions for nurses (end users)	Prompt questions for cocreator nurses	Prompt questions for managers
Attitude	• What are your impressions regarding the fact that a coworker from your department contributed to this nursing technological innovation? • What is your opinion on dedicating time to enable nurses to create and execute nursing technological innovations for nursing care?	• Can you elaborate on how the cocreation process began and how it evolved? • What is your opinion on dedicating time to enable nurses to create and execute nursing technological innovations for nursing care?	• What are your thoughts on the fact that a nurse from the nursing department has made a substantial contribution to this nursing technological innovation? • What is your opinion on dedicating time to enable nurses to create and execute nursing technological innovations for nursing care?
Acceptance	• What were your initial impressions of the nursing technological innovation when you first began working with it?	• What were your initial impressions of the nursing technological innovation when you first began working with it?	• What responses did you observe from nurses when they were introduced to the nursing technological innovation for the first time?
Adoption	• In what ways do you integrate the nursing technological innovation into your daily nursing tasks and responsibilities?	• In what ways do you integrate the nursing technological innovation into your daily nursing tasks and responsibilities?	• What are the viewpoints within the department regarding the utilization of the nursing technological innovation?
Usability	• What is your experience with the nursing technological innovation at work?	• What is your experience with the nursing technological innovation at work?	• What are the experiences of the nurses regarding the use of the nursing technological innovation?
Nursing practice	• What transformations do you observe in nursing practice as a result of the application of the nursing technological innovation?	• What transformations do you observe in nursing practice as a result of the application of the nursing technological innovation?	• What transformations do you observe in nursing practice as a result of the application of the nursing technological innovation?
Patient care	• What changes do you perceive in patient outcomes as a result of the application of the nursing technological innovation?	• What changes do you perceive in patient outcomes as a result of the application of the nursing technological innovation?	• What changes do you perceive in patient outcomes as a result of the application of the nursing technological innovation?

Data Analysis

The semistructured interviews were transcribed verbatim [19]. A 6-step thematic analysis approach (Table 2) was applied using ATLAS.ti software (version 23; Lumivero) [26]. Open coding and constant comparison techniques were used during the analysis [19]. An iterative process was used between data collection and data analysis. Refining the interview questions made the collection of the data more focused and specific as the process developed [19]. Audio recordings were transcribed, read, and reread by SvS in order to become familiar with the data. Data were initially coded using open coding by SvS. The 1st, 3rd, 6th, and 10th transcripts were coded separately, both by SvS and TvH. The initial codes of SvS and TvH were compared and discussed until consensus was reached. Main themes were generated through an iterative process, including constant comparison, and further elaborated and finalized by SvS, TvH, and OH. A codebook was used for insights into coding and code development [19].

Table 2. Thematic analysis.

Phase	Explanation
Phase 1: familiarization with the data	SvS conducted interviews with the participants and transcribed them. Transcripts were read and reread by SvS to familiarize herself with the data and obtain an overall impression.
Phase 2: generation of initial codes	Data were extracted from the transcripts and initially coded using open coding by SvS. The 1st, 3rd, 6th, and 10th transcripts were coded separately, both by SvS and TvH. The initial codes of SvS and TvH were compared and discussed until consensus was reached. The memos were processed systematically. Codes remained linked to the transcriptions (quotes) and the data extracts, using ATLAS.ti (Lumivero).
Phase 3: broader levels of themes	SvS developed broad potential main themes by comparing the initial codes. During meetings with SvS, TvH, and OH, consensus was reached, resulting in potential main themes derived from the data. During follow-up meetings with SvS, TvH, and OH, a description of the potential main themes was created and deliberated upon.
Phase 4: reviewing and refining the candidate themes	SvS compared the potential main themes with the coded data extracts and the raw data (transcriptions), discussing inconsistencies with TvH and OH. This process led to the identification of the main themes, which were visualized with a thematic map. The themes were refined by SvS and peer reviewed by TvH and OH.
Phase 5: defining and naming themes	SvS revisited the data extracts to define the content of each main theme. In meetings, SvS, TvH, and OH discussed the essence of each theme and developed the overall story line, resulting in the final naming of the main themes.
Phase 6: producing the report	SvS drafted the report and included detailed examples and quotes supporting the main themes. TvH and OH provided feedback, and adjustments were made accordingly. Final recommendations were implemented in the report.

In addition to the thematic analysis, we performed a secondary analysis using the SWOT matrix. This secondary analysis was carried out to make the data more accessible for the readers and support strategic changes within an organization. The SWOT analysis has been shown to be a valid method in previous studies for facilitating data interpretation for readers and informing strategic planning and decision-making in practice [27,28]. In several meetings, detailed examples of the results, with related quotes of the main themes, were discussed and analyzed and subsequently linked to a component of the SWOT matrix.

Rigor and Reflexivity

Different techniques were used to enhance trustworthiness [19]. The credibility of this study was established by the investigators by creating a nonjudgmental ambiance during the semistructured interviews to obtain the participants' perspective. The chance for bias was reduced by transcribing the content of the interviews verbatim. Member checking was used to ensure the accuracy of the transcripts, with participants reviewing the transcripts to validate their thoughts and ideas [19]. The performed member check resulted in minimal textual changes to the data.

Memos were reported during data collection and analysis to support the investigation process and the thoughts that had occurred to the investigator. The conformability of the interpretation and credibility of the data were both enhanced by investigator triangulation during the data analysis and peer feedback during researcher meetings, including the investigators SvS, TvH, OH, and HSMK. Due to peer feedback, potential meanings and a wider range of perspectives were revealed. Reflexivity, ensured by the investigator's critical view of the interview process and comments from other investigators, increased depth and improved accuracy.

A thick description was pursued to ensure the transferability of this study through the diversity of the population base, the number of participants, and the length of the interviews for imitability. To enhance the reflexivity and transparency of the investigators, a logbook was kept, detailing all changes and decisions made throughout the process of this study [19].

Ethical Considerations

Throughout this study, the principles of the Declaration of Helsinki and the code of conduct for health research were followed [29,30]. The study was examined and approved by the medical ethics review committee of the Erasmus Medical Center for a nonmedical scientific research application (MEC-2023-0021). Participants were informed and invited to the study by OH. They received the investigator's email and had 1 week to express interest. Nonresponders received a reminder. The investigator sent a participant information letter and informed consent upon contact. Participants could ask further questions via email. Interviews were scheduled after participants confirmed their approval. Before the interview, they provided written informed consent. The participants did not receive any compensation for their participation in the study. During the study, data collection was conducted with confidentiality, and data were processed anonymously according to the European Union General Data Protection Regulation [29]. Only the investigators SvS, TvH, OH, and HSMK had access to the source data.

Results

Characteristics of the Participants

The duration of the interviews ranged from 26 to 62 minutes, with an average duration of 44 (SD 10.8) minutes. Participants' ages ranged from 20 to 67 years. Various departments, including the intensive care unit, children's ward, and the recovery department, were represented in the study. Table 3 presents detailed demographic information, including gender, age range, department, adoption rate, number of nurses participating as cocreators and end users, and the number of health care managers.

Table 3. Demographic data of the participants (N=15).

Characteristics	Participants, n (%)
Gender	
Women	10 (67)
Men	5 (33)
Age range (y)	
20-29	1 (7)
30-39	5 (33)
40-49	3 (20)
50-59	3 (20)
60-67	3 (20)
Department	
ICU[a] (adults)	7 (47)
ICU (thoracic surgery and cardiac monitoring)	2 (13)
Center for home ventilator and respiratory disorders in children	2 (13)
ICU (children)	1 (7)
Recovery	1 (7)
Day treatment (children)	2 (13)
Adoption rate[b]	
Innovator[c]	10 (67)
Early adopters[d]	3 (20)
Early majority[e]	2 (13)
Late majority[f]	0 (0)
Laggards[g]	0 (0)
Profession	
Nurse (cocreator)	6 (40)
Nurse (end user)	5 (33)
Health care manager	4 (27)

[a]ICU: intensive care unit.

[b]Adoption rate was measured by asking the participants which category they identified with.

[c]Driven by change and introduction of innovations.

[d]Leading in adopting innovations.

[e]Deliberately adopting innovations.

[f]Adoption is an economic necessity and a response to social peer pressure.

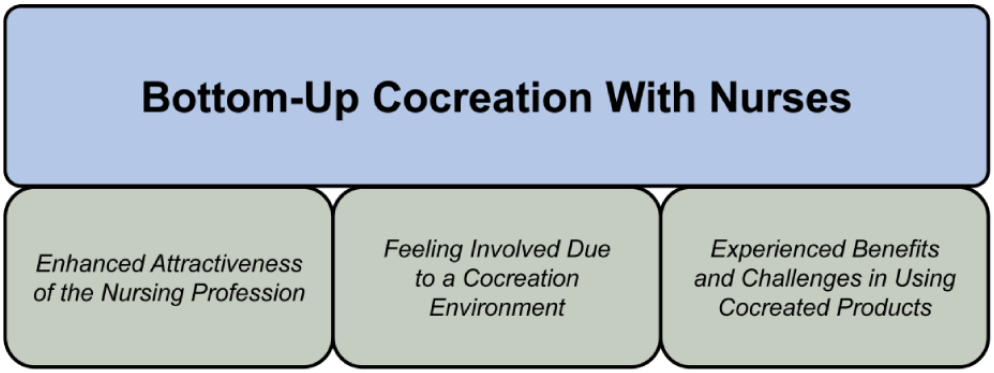
[g]Laggards have a traditional view and are more skeptical about innovations [26].

Bottom-Up Cocreation With Nurses

Overview

The thematic analysis uncovered the following three main themes concerning the participants' experiences with the bottom-up cocreation approach and resulting nursing technological innovations: (1) enhanced attractiveness of the nursing profession, (2) feeling involved due to a cocreation environment, and (3) experienced benefits and challenges in using cocreated products. Figure 1 provides a visual representation of the main themes.

Figure 1. Thematic map with the 3 main themes.



Enhanced Attractiveness of the Nursing Profession

All participants experienced that the bottom-up cocreation approach with nurses contributed to the overall attractiveness of the nursing profession. The nursing profession's attractiveness was enhanced as participants were enthusiastic about participating in cocreation, finding it enjoyable and engaging. Moreover, some participants felt that they were taken more seriously as nurses as they were the ones who actually experienced the problems in practice and were given an opportunity to use their expertise to address these challenges and drive improvements in nursing practice:

> *It makes work more enjoyable. You feel taken seriously as a nurse. You can share your ideas. You can use your expertise to improve something. So I think that's a lot of fun and very inspiring, and it also leads to something useful in the end.* [P11, cocreator]

Some participants emphasized the importance of having affinity with cocreation. Most participants felt a sense of pride in having a cocreation department within the organization and being a part of a cocreation process. A few participants expressed modesty about contributing to the cocreation product:

> *Well, I think we were genuinely proud to be able to offer the product to the patient. It was a joint effort and of course it was also very important for us that our name was set out on the product. So another team could not say, "Look at this!" It really makes you feel good if you, if you can make something that you really like to use, then that is actually really nice.* [P8, end user]

In addition, nurses were more attracted to the profession because they assumed multiple roles during cocreation. They acted as developers among designers, students, and the nursing department, actively contributing ideas and promoting products within the departments. Some participants expressed that the opportunity to become involved in cocreation made their work more interesting, and, due to this, they did not look for employment elsewhere. Cocreation with nurses was also inspiring for some participants. It motivated other nurses to address practical problems encountered during their work and explore possibilities for solving them:

> *But you're then involved in care in just a slightly different way than in standard care. And that also gives new energy and new ideas. And maybe other colleagues could also get new ideas and think, well, I have seen it a few times now, so what if I can come up with some ideas like this. Then they might think, oh well, perhaps this is something that we can also work on. This could also inspire others.* [P13, cocreator]

Feeling Involved Due to a Cocreation Environment

All participants reported enhanced collaboration through bottom-up cocreation involving various stakeholders. Collaboration included clear communication and teamwork among nurses, designers, and technically oriented students. Most participants mentioned that collaboration within the organization increased their willingness to adopt certain products that were aligned with nursing practice. One participant preferred internal product development but acknowledged that external innovations could still benefit end users:

> *If you get everyone involved, then you will have much more support and that is also the case with new things. I have experienced a few times in my life that we were getting an email on Sunday night. "From Monday we will do it like this." Well, then nurses will be digging in their heels. Whereas if you say, we're working on this, and who wants to contribute their views? Then you create support; much more people will then be willing to take it from there.* [P9, cocreator]

> *I like it when the product comes from us, I like that, but it does not necessarily have to come from us.* [P8, end user]

Some participants highlighted that collaboration among nursing departments, other hospitals, and organizations contributed to knowledge exchange, product distribution, and external production. However, one participant mentioned poor communication among nursing departments concerning new developments within the organization:

> *But we, we do note things, and we share them. It's important that you know this. We're not all working in separate groups.... On the contrary, a lot of information is shared among the four of us. Because we have to intervene and know what the other thinks.* [P10, manager]

> *No, that is also very poorly communicated. And the same applies to these products.* [P9, cocreator]

Some participants mentioned encountering practical problems but found it challenging to address these problems without the support of an innovation department with a cocreation group. The presence of such a department allowed the active participation of nurses and patients in creating specific products. Furthermore, a subgroup of participants engaged in cocreation with both patients and their parents, resulting in products tailored to their specific desires and needs. Patients were enthusiastic and felt heard, leading to the active use of end-user products:

> *Yes, parents have been involved in the co-creation process. That is, in the entire process of designing, they were not just allowed to give input, but were also able make direct choices, about what benefits them the most.* [P12, manager]

An aspect experienced by most participants was the need for support during the cocreation process. Several participants emphasized the importance of a supportive manager in facilitating and allocating dedicated time for the cocreation process. They noted that cocreation was unlikely to occur without managerial backing. Participants reported that supportive managers who recognized the value of cocreation and its potential to save time in nursing care were more proactive in addressing and resolving practical problems:

> *Managers can stimulate it. I think that people working in healthcare can be very creative, but often think, what shall I say, in boxes, because it's not possible, because we're too busy. But actually that is such a shame, because we just have a lot to offer and we have the knowledge and experience to be able to create something beautiful. And I think if a manager who (A) stimulates and (B) gives the space that is required, this will ensure that people are more motivated to contribute. Because they're aware that improvement is possible.* [P11, cocreator]

Experienced Benefits and Challenges in Using Cocreated Products

Benefits

Participants found the products resulting from the bottom-up approach to be highly practical and user-friendly when used during their nursing tasks. All participants expressed immediate positivity toward the product due to its straightforward design. They could use the product without needing special training. Nurses also immediately experienced positive outcomes during their tasks. For instance, one participant mentioned that a newly created clamp for placing pressure transducers at the correct height on infusion poles improved nursing efficiency. They also reported that this clamp alleviated wrist complaints:

> *It is faster, yes, you press it and put it on the pole, loosen it and put it back on the transport cart.* [P6 end user]

Some participants found that the products improved their ability to organize and manage nursing tasks. For example, a product named KoosGuard contributed to an improved organization of patient monitoring wires, providing enhanced structure during nursing care. Another product, the cannula emergency bag, ensured that nurses, patients, and parents had all the materials neatly arranged in case of an emergency, enhancing preparedness and efficiency:

> *Overview yes, you immediately have a clear overview. Because it's easy to open it, everything is in place and it's compact. And because it was developed together with the parents and the nurses, of course, they really like that. It fits their needs.* [P12, manager]

Several participants emphasized the importance of the products' quality and safety. Participants highlighted that the professional appearance of the product and its suitability for nursing practice served as indicators of quality. Participants reported that these well-suited products reduced the need for makeshift solutions, resulting in fewer inconvenient situations during their work:

> *Yes, the well-known example is that we often use band-aids and things, but that is to help you out for the time being. But practice has already shown that there are much more convenient solutions.* [P11, cocreator]

Certain participants emphasized the role of the products in enhancing patient safety. Participants noted that a specific product effectively prevented infusion line entanglement and tension, thus reducing the risk of accidental removal. In addition, participants mentioned a product designed to decrease the risk of postsurgery wound contamination, promoting improved healing outcomes:

> *You really want those lines to be properly arranged. To make sure that not one of them is hidden from your view, because if one of those lines is pulled, the infusion may accidentally get disconnected or the patient can accidentally end up lying on top of that line because you can't see it.* [P14, manager]

Challenges

In addition to the benefits of the products, participants also experienced some challenges in their use. Some participants mentioned that some product designs had flaws, which hindered their effectiveness. For example, a participant heard from some patients that they still thought that the cannula emergency bag was too large to be used conveniently, and, due to this, it was no longer used. In addition, it was indicated that some products looked disposable, so nurses threw them away:

> *The only negative feedback came from a patient. She just thought the bag was too big.* [P8, end user]

A few participants pointed out that some products were prone to getting lost or becoming untraceable. It was emphasized that these products should be readily available to nurses at designated and standardized locations. Otherwise, the product will be used less or not used at all. Furthermore, some participants shared their struggle with making changes, leading to a heightened risk of reverting to old routines:

> *Then it takes a lot of time anyway. And you think: well, I'll just do it in the old-fashioned way, using a gauze.* [P3, end user]

SWOT Analysis Findings

Detailed examples of the main themes were divided into the 4 components of the SWOT matrix. Multimedia Appendix 1 presents an overview of all the main themes, along with detailed examples, illustrative quotations, and the corresponding SWOT categorization. The “strengths” and “weaknesses” encompass the experienced advantages and disadvantages associated with the bottom-up cocreation approach and the products developed through this approach. The “opportunities” and “threats” are examples that could be associated with potential future opportunities and challenges of the bottom-up cocreation approach and developed products. Most of the examples were connected to the components “strengths” and “opportunities.”

Discussion

Principal Findings

In this study, we sought to explore hospital nurses’ experience with a bottom-up cocreation approach and the resulting nursing technological innovations developed through this approach. Notably, through a SWOT matrix, we mainly revealed positive experiences of participants cocreating nursing technological innovations, with several strengths and opportunities. Participants’ experiences can be described in 3 themes: enhanced attractiveness of the nursing profession, feeling involved due to a cocreation environment, and experienced benefits and challenges in using cocreated products. The findings indicate that participants found the nursing technological innovations easily applicable in nursing practice due to their user-friendly design, the ability to enhance nursing efficiency, patient safety, and time-saving potential during nursing work. A collaborative culture may lead to higher work satisfaction among nurses, inspire cocreation, and promote nursing pride.

Comparison With Prior Work

This study shows how cocreating nursing technological innovations positively impacts the attractiveness of the nursing profession. The attractiveness of the nursing profession was enhanced by nurses finding joy in their roles, feeling a sense of pride, and being taken seriously in their work. This aligns with findings from a cross-sectional study, which emphasizes the significance of decision-making involvement in health care services, such as equipment, technologies, and processes for employment preferences of nursing students [31]. In addition, it aligns with a systematic review indicating that an increase in career opportunities and challenges positively impacts nurse retention. It reduces nurses’ desire to leave a particular workplace [32].

Furthermore, this study emphasizes several collaboration advances due to a cocreation environment within the organization, such as clear communication and teamwork among nurses, designers, and technically oriented students. Moreover, a cocreation environment within the organization contributed to knowledge exchange, product distribution, and external production. An in-depth case study [14] reinforced the importance of an innovative environment within the organization. It could increase the diffusion of nurse-led innovations within and outside the organization. Particularly, managers seem to have an important role in collaborating and communicating with each other about innovations and supporting working methods in nursing departments. Existing literature [33,34] reinforces the notion that supportive leadership is crucial for fostering a culture of innovation and enhancing the sustainability and impact of health care innovation settings.

In addition, this study shows that innovations cocreated with nurses led to multiple benefits for nursing practice. Cocreated nursing technological innovations were described as user-friendly, highly practical, and easy to integrate into nursing tasks, consistent with findings from a quasi-experimental study [35]. In this quasi-experimental study, trained innovative nurses were able to identify issues in nursing practice and apply creative thinking strategies to create innovative products that matched their nursing work routines. Furthermore, the cocreated nursing technological innovations met nursing quality requirements and reduced the need for improvised solutions, thereby minimizing inconvenient situations in nursing practice. Patient safety was also enhanced by the cocreated products, which prevented issues such as infusion line entanglement and reduced the risk of wound contamination. The incubator traffic light is an example from the literature of a cocreated nursing technological innovation that meets quality standards and enhances patient safety. The incubator traffic light features a visual feedback system for neonatal incubators designed to enhance hand hygiene compliance [1,2]. Incubators with the incubator traffic light visual feedback system demonstrated significantly higher compliance with the correct drying times compared to those without this feature. This could potentially enhance patient safety by reducing infection rates in neonatal intensive care units [1]. Although the cocreated nursing technological innovations offered several advantages, the study also encountered certain challenges in their use, for example, flaws in the design of the innovations, which hindered their effectiveness or ability to make changes in the nursing routines. This study shows that changes in nursing processes may increase the risk of reverting to old routines, which aligns with findings from a cross-sectional survey that indicated how changes in nursing work processes could influence established work habits [36].

While this study did not include patients or parents directly, the participants acknowledged the significance of their perspective during the innovation process. It was mentioned that some of the cocreated innovations were developed in collaboration with patients and parents, resulting in innovations aligning with their expectations and needs. This is consistent with existing literature, where patients, family caregivers, and clinicians cocreated a mobile health app for heart failure self-management tailored to the specific health care context [37]. Furthermore, it matches with a qualitative co-design article including family, physicians, researchers, patients, and industry partners. The study revealed important differences between the participants’ preference for functional requirements of a mobile health app [38].

Limitations

Limitations of the study include its single-site focus on a specific academic hospital and limited transferability to other health

care settings [19]. Furthermore, there may be a sampling bias, as interested individuals were required to actively contact the investigator via email to participate in this study. These individuals might have been more likely than other professionals to embrace the innovations and encourage their adoption within their teams. In addition, all the participants scored themselves above average on the adoption rate scale (Table 3), and 10 (67%) of the 15 participants considered themselves to be "innovators" [25]. Literature shows that only 2.5% of the population belongs to the group of innovators [25]. This suggests a potential bias in the participant group. Including nurses as cocreators may have introduced selection bias, as their enthusiasm and assertiveness toward innovations may have influenced the results in a more positive manner [39]. Finally, the sample size consisted of 6 (40%) nurses as cocreators, 5 (33%) nurses as end users, and 4 (27%) managers. The limited insights from the managers' perspective may affect the comprehensiveness of the findings in the study [19].

Implications and Further Research

The findings of this study suggest several implications. Health care organizations could consider adopting the bottom-up cocreation approach with nurses as it has the potential to enhance nurses' work satisfaction, a sense of pride among nurses, and adoptable nursing technological innovations in nursing care. Furthermore, cocreation can contribute to accessible collaborations between stakeholders inside and outside the organization. In addition, cocreation with nurses has the potential to facilitate the development of innovations that enhance the efficiency of nursing work. Simple, professional, and time-saving innovations created through this approach could improve nursing practices.

Further research is recommended to explore the impact of the bottom-up cocreation approach and the resulting nursing technological innovations in other health care centers, aiming to improve the transferability of this way of working. Moreover, to better address patients' expectations and needs during the cocreation process and resulting innovations, further research is recommended to investigate their experiences throughout the cocreation process and its outcomes.

Conclusions

This study highlights substantial positive experiences of cocreating nursing technological innovations in nursing care. The findings underscore that cocreation with nurses enhances the appeal of the nursing profession. Participants perceived cocreation as enjoyable, leading to heightened work satisfaction and a sense of pride in their nursing role. Moreover, cocreation with nurses fosters the development of nursing technological innovations that align with nursing practice challenges, thereby facilitating their adoption within nursing care. Ultimately, embracing a culture of cocreation has the potential to foster ongoing improvements and innovations in nursing care, further contributing to the professionalization of the field.

Acknowledgments

The authors extend their appreciation to Myrthe van der Zanden, MSc, nursing scientist, for reviewing the interview guide and participating in the pilot interview.

Data Availability

The datasets generated or analyzed during this study are not publicly available due the participants involved in this research did not authorize the public dissemination of their data through written consent.

Authors' Contributions

SvS contributed to conceptualization, writing the original draft, data curation, formal analysis, investigation, methodology, project administration, and visualization. OH contributed to recruitment, conceptualization, reviewing and editing the manuscript, formal analysis, investigation, methodology, project administration, and supervision. HSMK contributed to conceptualization, methodology, and reviewing and editing the manuscript. TvH contributed to conceptualization, reviewing and editing the manuscript, data curation, formal analysis, investigation, methodology, project administration, and supervision.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Main themes; detailed examples; illustrative quotes; and strengths, weaknesses, opportunities, and threats categorization.
[DOCX File , 23 KB - humanfactors_v12i1e60543_app1.docx]

References

1. van Gils RH, Helder OK, Wauben LS. Incubator traffic light: the development of an alcohol-based hand rub dispenser system for neonatal incubators with visual feedback to improve hand hygiene compliance. BMJ Innov 2019 Sep 06;5(2-3):70-77. [doi: 10.1136/bmjinnov-2018-000301]

2. van Gils RH, Kornelisse RF, Dankelman J, Helder OK. Validation of a hand hygiene visual feedback system to improve compliance with drying time of alcohol-based hand rub in a neonatal intensive care unit: the incubator traffic light system. J Hosp Infect 2024 Mar;145:210-217 [FREE Full text] [doi: 10.1016/j.jhin.2024.01.007] [Medline: 38272126]
3. van der Zanden M, Helder O, Westland H, Ista E. Barriers and facilitators for the implementation of co-created nursing innovations in a Dutch university hospital. JBI Evid Implement 2024 Jun 28;22:11. [doi: 10.1097/XEB.0000000000000444] [Medline: 38932713]
4. Rudnicka E, Napierała P, Podfigurna A, Męczekalski B, Smolarczyk R, Grymowicz M. The World Health Organization (WHO) approach to healthy ageing. Maturitas 2020 Sep;139:6-11 [FREE Full text] [doi: 10.1016/j.maturitas.2020.05.018] [Medline: 32747042]
5. Dugstad J, Eide T, Nilsen ER, Eide H. Towards successful digital transformation through co-creation: a longitudinal study of a four-year implementation of digital monitoring technology in residential care for persons with dementia. BMC Health Serv Res 2019 Jun 10;19(1):366 [FREE Full text] [doi: 10.1186/s12913-019-4191-1] [Medline: 31182093]
6. de Jong JP, Rigtering C, Spaans L. Heroes of diffusion: making user innovations widely available. Res Policy 2023 Oct;52(8):104840. [doi: 10.1016/j.respol.2023.104840]
7. Wang S, Bolling K, Mao W, Reichstadt J, Jeste D, Kim HC, et al. Technology to support aging in place: older adults' perspectives. Healthcare (Basel) 2019 Apr 10;7(2):60 [FREE Full text] [doi: 10.3390/healthcare7020060] [Medline: 30974780]
8. Patey AM, Soong C. Top-down and bottom-up approaches to low-value care. BMJ Qual Saf 2023 Feb 14;32(2):65-68. [doi: 10.1136/bmjqs-2022-014977] [Medline: 36517225]
9. Kernebeck S, Jux C, Busse TS, Meyer D, Dreier LA, Zenz D, et al. Participatory design of a medication module in an electronic medical record for paediatric palliative care: a think-aloud approach with nurses and physicians. Children (Basel) 2022 Jan 06;9(1):82 [FREE Full text] [doi: 10.3390/children9010082] [Medline: 35053707]
10. Kernebeck S, Busse TS, Jux C, Dreier LA, Meyer D, Zenz D, et al. Evaluation of an electronic medical record module for nursing documentation in paediatric palliative care: involvement of nurses with a think-aloud approach. Int J Environ Res Public Health 2022 Mar 18;19(6):3637 [FREE Full text] [doi: 10.3390/ijerph19063637] [Medline: 35329323]
11. Grynne A, Browall M, Fristedt S, Ahlberg K, Smith F. Integrating perspectives of patients, healthcare professionals, system developers and academics in the co-design of a digital information tool. PLoS One 2021;16(7):e0253448 [FREE Full text] [doi: 10.1371/journal.pone.0253448] [Medline: 34270564]
12. Bryant L, Sedlarevic N, Stubbs P, Bailey B, Nguyen V, Bluff A, et al. Collaborative co-design and evaluation of an immersive virtual reality application prototype for communication rehabilitation (DISCOVR prototype). Disabil Rehabil Assist Technol 2024 Jan;19(1):90-99. [doi: 10.1080/17483107.2022.2063423] [Medline: 35442823]
13. Prahalad C, Ramaswamy V. Co-creation experiences: the next practice in value creation. J Interact Mark 2004 Aug 01;18(3):5-14. [doi: 10.1002/dir.20015]
14. Rigtering C, Spaans LJ, de Jong JP. How to bridge the nurse innovation-diffusion gap? An in-depth case study of Create4Care. Front Public Health 2023;11:1209965 [FREE Full text] [doi: 10.3389/fpubh.2023.1209965] [Medline: 37601181]
15. McDermid F, Mannix J, Peters K. Factors contributing to high turnover rates of emergency nurses: a review of the literature. Aust Crit Care 2020 Jul;33(4):390-396. [doi: 10.1016/j.aucc.2019.09.002] [Medline: 31836449]
16. Racy S, Davidson PM, Peeler A, Hager DN, Street L, Koirala B. A review of inpatient nursing workload measures. J Clin Nurs 2021 Jul;30(13-14):1799-1809. [doi: 10.1111/jocn.15676] [Medline: 33503306]
17. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 09;6:3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]
18. Devi DH, Duraisamy K, Armghan A, Alsharari M, Aliqab K, Sorathiya V, et al. 5G technology in healthcare and wearable devices: a review. Sensors (Basel) 2023 Feb 24;23(5):2519 [FREE Full text] [doi: 10.3390/s23052519] [Medline: 36904721]
19. Holloway I, Galvin L. Qualitative Research in Nursing and Healthcare, Fourth Edition. Hoboken, NJ: John Wiley & Sons; 2016.
20. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
21. Boeije HR. Analysis in Qualitative Research. Thousand Oaks, CA: SAGE Publications; 2010.
22. Creswell JW, Poth CN. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Thousand Oaks, CA: SAGE Publications; 2024.
23. Brandsma T, Stoffers J, Schrijver I. Advanced technology use by care professionals. Int J Environ Res Public Health 2020 Jan 23;17(3):742 [FREE Full text] [doi: 10.3390/ijerph17030742] [Medline: 31979307]
24. Farley HL, Baumlin KM, Hamedani AG, Cheung DS, Edwards MR, Fuller DC, et al. Quality and safety implications of emergency department information systems. Ann Emerg Med 2013 Oct;62(4):399-407 [FREE Full text] [doi: 10.1016/j.annemergmed.2013.05.019] [Medline: 23796627]
25. Rogers EM. Diffusion of Innovations, Fourth Edition. New York, NY: Free Press; 2010.

26. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
27. Baez-Leon C, Palacios-Ceña D, Fernandez-de-Las-Peñas C, Velarde-García JF, Rodríguez-Martínez M, Arribas-Cobo P. A qualitative study on a novel peer collaboration care programme during the first COVID-19 outbreak: a SWOT analysis. Nurs Open 2022 Jan;9(1):765-774 [FREE Full text] [doi: 10.1002/nop2.1128] [Medline: 34773372]
28. Topuz Ş, Yilmaz Sezer N, Aker MN, Gönenç IM, Öner Cengiz H, Er Korucu A. A SWOT analysis of the opinions of midwifery students about distance education during the Covid-19 pandemic a qualitative study. Midwifery 2021 Dec;103:103161 [FREE Full text] [doi: 10.1016/j.midw.2021.103161] [Medline: 34628182]
29. Regulation (EU) 2016/679 of the European parliament and of the council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union. 2016 May 4. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2025-03-15]
30. Code of conduct for health research. Central Committee on Research Involving Human Subjects. URL: https://english.ccmo.nl/investigators/legal-framework-for-medical-scientific-research/codes-of-conduct/code-of-conduct-for-health-research [accessed 2024-03-25]
31. Terry D, Peck B, Baker E, Schmitz D. The rural nursing workforce hierarchy of needs: decision-making concerning future rural healthcare employment. Healthcare (Basel) 2021 Sep 18;9(9):1232 [FREE Full text] [doi: 10.3390/healthcare9091232] [Medline: 34575006]
32. de Vries N, Boone A, Godderis L, Bouman J, Szemik S, Matranga D, et al. The race to retain healthcare workers: a systematic review on factors that impact retention of nurses and physicians in hospitals. Inquiry 2023;60:469580231159318 [FREE Full text] [doi: 10.1177/00469580231159318] [Medline: 36912131]
33. Albert NM. Operationalizing a nursing innovation center within a health care system. Nurs Adm Q 2018;42(1):43-53. [doi: 10.1097/NAQ.0000000000000266] [Medline: 29194332]
34. Barr TL, Malloch K, Ackerman MH, Raderstorf T, Melnyk BM. A blueprint for nursing innovation centers. Nurs Outlook 2021;69(6):969-981 [FREE Full text] [doi: 10.1016/j.outlook.2021.05.006] [Medline: 34183188]
35. Gao L, Lu Q, Hou X, Ou J, Wang M. Effectiveness of a nursing innovation workshop at enhancing nurses' innovation abilities: a quasi-experimental study. Nurs Open 2022 Jan;9(1):418-427 [FREE Full text] [doi: 10.1002/nop2.1080] [Medline: 34687153]
36. Alsohime F, Temsah MH, Al-Eyadhy A, Bashiri FA, Househ M, Jamal A, et al. Satisfaction and perceived usefulness with newly-implemented electronic health records system among pediatricians at a university hospital. Comput Methods Programs Biomed 2019 Feb;169:51-57. [doi: 10.1016/j.cmpb.2018.12.026] [Medline: 30638591]
37. Woods L, Duff J, Roehrer E, Walker K, Cummings E. Design of a consumer mobile health app for heart failure: findings from the nurse-led co-design of Care4myHeart. JMIR Nurs 2019;2(1):e14633. [doi: 10.2196/14633] [Medline: 34345774]
38. Mrklas KJ, Barber T, Campbell-Scherer D, Green LA, Li LC, Marlett N, et al. Co-design in the development of a mobile health app for the management of knee osteoarthritis by patients and physicians: qualitative study. JMIR Mhealth Uhealth 2020 Jul 10;8(7):e17893 [FREE Full text] [doi: 10.2196/17893] [Medline: 32673245]
39. Bouter LM, Zielhuis GA, Zeegers MP. Textbook of Epidemiology. Houten, the Netherlands: Bohn Stafleu van Loghum; 2017.

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
SWOT: strengths, weaknesses, opportunities, and threats

Edited by A Kushniruk; submitted 06.06.24; peer-reviewed by E Huber, GC Nickel; comments to author 14.10.24; revised version received 05.12.24; accepted 10.02.25; published 31.03.25.

Please cite as:
van Steenis S, Helder O, Kort HSM, van Houwelingen T
Impact of Bottom-Up Cocreation of Nursing Technological Innovations: Explorative Interview Study Among Hospital Nurses and Managers
JMIR Hum Factors 2025;12:e60543
URL: https://humanfactors.jmir.org/2025/1/e60543
doi:10.2196/60543
PMID:

Original Paper

Capturing Requirements for a Data Annotation Tool for Intensive Care: Experimental User-Centered Design Study

Marceli Wac[1,2], BSc; Raul Santos-Rodriguez[1], MSc, PhD; Chris McWilliams[1,2], BSc, PhD; Christopher Bourdeaux[2], BCHIR, MB, MA

[1]Faculty of Engineering, University of Bristol, Bristol, United Kingdom

[2]University Hospitals Bristol and Weston NHS Foundation Trust, Bristol, United Kingdom

Corresponding Author:
Marceli Wac, BSc
Faculty of Engineering
University of Bristol
Queen's Building
University Walk
Bristol, BS8 1TR
United Kingdom
Phone: 44 1173315830
Email: m.wac@bristol.ac.uk

Abstract

Background: Increasing use of computational methods in health care provides opportunities to address previously unsolvable problems. Machine learning techniques applied to routinely collected data can enhance clinical tools and improve patient outcomes, but their effective deployment comes with significant challenges. While some tasks can be addressed by training machine learning models directly on the collected data, more complex problems require additional input in the form of data annotations. Data annotation is a complex and time-consuming problem that requires domain expertise and frequently, technical proficiency. With clinicians' time being an extremely limited resource, existing tools fail to provide an effective workflow for deployment in health care.

Objective: This paper investigates the approach of intensive care unit staff to the task of data annotation. Specifically, it aims to (1) understand how clinicians approach data annotation and (2) capture the requirements for a digital annotation tool for the health care setting.

Methods: We conducted an experimental activity involving annotation of the printed excerpts of real time-series admission data with 7 intensive care unit clinicians. Each participant annotated an identical set of admissions with the periods of weaning from mechanical ventilation during a single 45-minute workshop. Participants were observed during task completion and their actions were analyzed within Norman's Interaction Cycle model to identify the software requirements.

Results: Clinicians followed a cyclic process of investigation, annotation, data reevaluation, and label refinement. Variety of techniques were used to investigate data and create annotations. We identified 11 requirements for the digital tool across 4 domains: annotation of individual admissions (n=5), semiautomated annotation (n=3), operational constraints (n=2), and use of labels in machine learning (n=1).

Conclusions: Effective data annotation in a clinical setting relies on flexibility in analysis and label creation and workflow continuity across multiple admissions. There is a need to ensure a seamless transition between data investigation, annotation, and refinement of the labels.

(JMIR Hum Factors 2025;12:e56880) doi:10.2196/56880

KEYWORDS

ICU; intensive care; machine learning; data annotation; data labeling; annotation software; capturing software requirements

Introduction

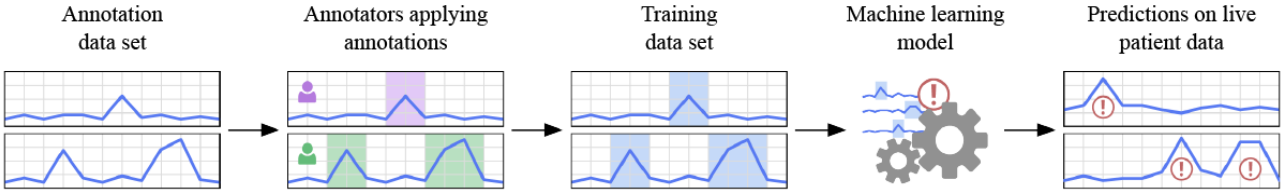
Background

Artificial intelligence (AI) is a field concerned with leveraging computers to mimic human cognitive functions, such as problem-solving and decision-making [1,2]. Depending on the task, this process can use a variety of methods and take many forms ranging from relatively simple rule-based expert systems (following if-then pathways) through regression (modeling the relation between different variables to predict their values) and clustering (grouping objects with similar properties together) to more complex systems such as artificial neural networks (computing systems imitating the anatomy of human brains capable of modeling nonlinear processes) [1,3,4]. Particularly complex problems may require solutions that cannot be achieved by traditional approaches, such as preprograming the desired behavior. Instead, techniques such as machine learning (ML) form a subset of AI in which the algorithms are used to analyze large amounts of data to derive a method for computing a solution [5]. In ML, the practice of arriving at a solution is called "learning" (or training), and the produced output is known as the ML model. The heavy reliance on data in ML highlights the importance of ensuring the appropriate quantity and quality of data and signifies its impact on the effectiveness of the created model [6]. The continuously increasing popularity of ML and its rapid adoption rate in life sciences suggest that AI is at the forefront of bringing innovation to health care [7,8].

Intensive care units (ICUs) are busy and complex health care environments where critically ill patients frequently require continuous monitoring and multiple-organ support. To provide care for those patients, clinicians working in the ICUs use a broad range of medical devices, such as ventilators, monitoring devices, and intravenous pumps and lines among many others. The information captured by those devices, as well as that entered by the ICU staff, is collected and collated in a clinical information system, which enables health practitioners to access large quantities of data routinely required as part of their job. This data-rich nature and the direct influence of data on the provision of care provide a tremendous opportunity for the deployment of ML models in health care and in particular ICUs [9].

While the data gathered in the ICUs can often be used directly, for example, to present the correlation between different vital signs and the patient prognosis, complex tasks that rely heavily on clinical experience and expertise may require human involvement to provide further guidance [10]. This guidance, most frequently referred to as labels or annotations [11], can deliver additional information or context to the existing data. An example of such a label could be a "yes" or "no" indicator of whether a patient is ready for discharge, or the type and size of tumor present on a radiograph. With this further knowledge, ML models can take advantage of the human experience to tackle complex problems and deliver solutions that could not be inferred from raw data alone [12] (Figure 1).

Figure 1. Annotating data can provide additional information and context necessary to train machine learning models.



While the large volumes of data available in intensive care offer significant opportunities, working with these data also presents a unique set of challenges, as the effort required to annotate the dataset increases proportionally to its size [13]. Furthermore, the complex nature of the health care data and the fact that applying a single label can require clinicians to look at multiple parameters, patient history, and laboratory results make the annotation task highly labor-intensive. With the clinician's time being a remarkably valuable resource, ensuring the effectiveness of the data annotation workflow is paramount. The lack of an annotation system tailored to the unique nature of the health care data (eg, accessing a subset of relevant variables from a list of potentially hundreds of parameters [14]) further complicates this problem.

In addition to the difficulties associated with working within clinical settings, it is equally important to consider the inherent challenges of data annotation in their own right. These include the need to account for multiple different types of bias that could be introduced throughout the process [15]. For example, the annotators themselves may have their own preferences and preconceptions that will guide the approach they assume when annotating data. This could stem from the amount and diversity of their clinical experience in a context specific to the annotation task, such as the familiarity with the clinical problem, the specific population within which they have treated it, or even the specific tools and methods they have used in the past. Similarly, by annotating historical data that span the entire duration of the admissions, annotators may create labels with an inherent temporal or selection bias. Alternatively, annotators tasked with creating specific annotations may focus on a subset of the admission where they would expect to create the annotation, such as reliance on medication, which may decrease toward the end of the admission but could equally change throughout its overall course.

Current literature on the applications of computational methods within health care highlights data annotation as a primary bottleneck in the ML pipeline [16]. This effect is attributed to the need for domain-specific knowledge and therefore access the expert population, as well as the considerable investment of their time [16-18]. Because of this, existing methods aim to reduce the number of required annotations required for positive results by using a variety of techniques including active learning [17], machine-assisted annotation [19], or synthetic data generation [20]. Nevertheless, the majority of studies continue

to rely on human annotations in the cases where labels cannot be easily derived, or are inadequately documented in structured data [16]. Designing a tool for the annotation of large clinical datasets is therefore a problem that needs to be approached carefully. The expert nature of the annotators means that both their numbers and time are limited and are therefore resources that need to be used efficiently [16]. Furthermore, clinical datasets frequently aggregate data spanning several years, resulting in a volume of data that is infeasible for manual annotation, suggesting a need for a semiautomated approach that could scale up to an entire dataset with limited input from the domain experts.

Involving end users in the design process of new tools prior to their development is an important aspect of designing effective software [21]. It minimizes the risk of creating a system that is inefficient and helps ensure that the developed solution meets users' expectations and requirements [22,23]. In the context of the data annotation software used by experts in intensive care, this participatory design is especially critical, as each of the variety of roles (eg, junior doctors, doctors, nurses, and consultants) can generate a unique set of requirements. Furthermore, while the staff working in the ICUs are experts in the medical domain trained to treat patients, extracting their knowledge through data annotation is an entirely different process. For this reason, it is crucial to deploy a strategy that will ensure that the design of the tool follows a structured and systematic approach that can capture a wide variety of perspectives from its end users while also accounting for the needs and priorities of the data science experts who will use the created annotations.

Objectives

The primary objective of the research is to establish a set of criteria for the design of a data annotation tool that can be used effectively by clinicians to annotate time-series datasets from intensive care. To achieve this goal under the limitations of working within the health care setting, the system needs to account for limited access to the annotators and large volumes of data. Finally, to ensure the efficiency of the process, the subtleties of how clinicians approach the problem of data annotation need to be understood and accounted for in the design. The overarching goal is, therefore, to gather requirements for a data annotation platform that is purpose-built for the intensive care data and clinicians and, as such, one that facilitates an efficient annotation workflow in that domain.

Methods

Study Design

To understand how clinicians approach and reason about the data annotation process, we conducted an experimental study that involved members of the clinical staff from the ICU manually annotating excerpts of time-series data printed on paper. Observations taken during task completion served as a basis for analysis and were used to derive the requirements for the data annotation software. This choice of methodology was influenced by constraints associated with working within the clinical setting, such as the high demand and short supply of participants' time, and the direct relationship between the task completion and the requirements for the digital tool, which allowed for a natural and unobstructed data collection.

Participant Recruitment

Our participant recruitment followed a mixture of convenience and snowball sampling comprising an invitation email sent to the staff working in the ICU who had prior experience with mechanical ventilation treatment and signposting through the internal network at our research site—University Hospitals Bristol and Weston NHS Foundation Trust. This recruitment process yielded a sample size of 7 participants across 2 distinct job roles (6 junior doctors and 1 doctor), which provided limited feasibility for suitable analysis of the interrole differences in approach to annotation but was sufficiently rich for establishing the software requirements for the tool we set out to design. Comparative sample sizes can be observed in similar research in the field [24-26]. Before data collection, the annotation task was piloted with a clinical member of our team experienced in mechanical ventilation treatment, as well as 2 other members of our team from the ML background. The insights collected from the pilot activity allowed us to refine the set of clinical parameters used throughout the activity, devise the inclusion criteria for the underlying admissions, and diversify the data extracts to form a sample representative of a typical dataset. Our research site had a preestablished research partnership with the University of Bristol and a technologically enhanced ICU with widespread adoption of digital systems.

Theoretical Framework

To analyze the participants' actions performed during the simulated activity, we used a framework that served as a basis for characterizing the data annotation process and identifying the challenges and opportunities associated with performing the task. Our framework followed the Norman's Interaction Cycle (NIC) model which assumes that the interaction is a process of evaluation and execution between the user and the technology [27]. It outlines a 7-step process that begins at the goal and, through evaluation of available means and strategies, as well as the execution of these strategies, leads to achieving that goal in practice. The framework emphasizes the breakdown of the preproduction stage into planning (formulating the problem that needs to be solved), specification (outlining the strategies that can be used to tackle the problem), and performance (specifying actions that need to be taken to deploy the strategy). Furthermore, it highlights the importance of presenting the user with feedback over 3 separate stages including perception (tracking the outcome of performed actions), interpretation (analyzing the effect of the outcome), and comparison (evaluating whether their actions resulted in achieving the goal) [27]. Finally, the model captures the disparities between user intentions and actions permitted by the technology (gulf of execution) and the effort required to correctly interpret the results of their actions with regard to the desired outcome (gulf of evaluation) [27,28].

Data Collection

To facilitate data collection, we simulated a data annotation activity involving manual annotation of real intensive care patient data using pens and highlighter pens on excerpts of data

printed on paper sheets. The data were formatted to resemble the clinical information system present at the research site and displayed patient demographics in addition to a table with time-series data. Each column of the table corresponded to an hour of the day and each row corresponded to a specific parameter; the table's cells contained the readings of a parameter for a given hour (Figure 2). The outline brief for the activity prompted participants to create annotations, allowing them to create multiple annotations for a single admission. Crucially, no additional instructions were provided regarding the annotation format. Before the data collection, the activity was piloted internally with an experienced intensive care consultant (CB) and 2 ML experts (RSR and CMW), which helped us establish inclusion criteria for the selected admissions, broaden the characteristics of our admission sample, and refine the included parameters. The activity took place as a single in-person workshop, lasting approximately 45 minutes, during which each of the 7 participants annotated the same set of 5 unique admissions (Figure 3).

Participants were observed for the duration of task completion by a single observer who took field notes throughout the entire activity. These field notes captured the interactions of the participants with the annotation tasks and described the actions taken when annotating specific admissions, how multiple admissions were annotated, and questions asked by the participants during the task. The collected notes were then aggregated across the different stages of NIC and the annotated printouts were collected and used to supplement the analysis of the methods used by the participants to label the data.

Figure 2. The provided excerpts of time series data were formatted to resemble the interfaces of the clinical information system present at the research site (the depicted table contains data for illustrative purposes only and is trimmed for conciseness). FiO2: fraction of inspired oxygen; ICU: intensive care unit; PCWP: pulmonary capillary wedge pressure; PEEP: positive end-expiratory pressure.

Stay ID: 12341234 **Age:** 54 **Race:** WHITE **Admission time:** 14/8/2174 12:00 **Length of stay in hospital**

Visit in ICU during this admission: #1 **Gender:** F **Death time:** - **Discharge time:** 21/10/2174 14:00 **Length of stay in ICU:** 16

	8/10 07:00	08:00	09:00	10:00	11:00	12:00	13:
Vitals							
Heart Rate [bpm]	63	66	57	59	54	52	57
Arterial Blood Pressure Systolic [mmHg]	132	141	152	143	119	141	155
Arterial Blood Pressure Diastolic [mmHg]	65	67	73	73	69	72	76
Arterial Blood Pressure Mean [mmHg]	86	85	94	99	92	97	104
Non-invasive Blood Pressure Systolic [mmHg]			131	139	117	127	130
Non-invasive Blood Pressure Diastolic [mmHg]			75	81	70	75	72
Non-invasive Blood Pressure Mean [mmHg]			85	95	81	86	89
Cardiac Output (thermodilution) [L/min]							
Respiratory Rate [insp/min]	14	12	13	13	14	12	14
Respiratory							
PEEP Set [cm H2O]	5			5			
Temperature [C]	36.32						
PCWP [mmHg]							
FiO2 [%]	50			50			40

Figure 3. Participants annotated data in a variety of ways using pen and paper during the workshop held in the intensive care unit.

Task Selection

Capturing data labels in a manner representative of the general annotation process relied on selecting an appropriate task. On the one hand, the desired label could not be trivial and needed to pose a significant challenge, complex enough to prompt an in-depth analysis of the data. On the other hand, this task also needed to fall within the subset of domain knowledge that participants were familiar with and could therefore solve with a relative degree of confidence. Annotating weaning from mechanical ventilation satisfies both of these criteria due to its challenging nature and the relative bias of the person who performs it. The process of weaning can be characterized by great variability in practice [29], in which both the timing of when the weaning begins and the method in which it is delivered largely depend on the clinician in charge of the treatment [29,30]. Together with the lack of personalized guidelines [31] and a wide variety of ways in which patients can be weaned, mechanical ventilation weaning constitutes a label that is difficult to derive from data and requires domain expertise, which makes it suitable for this task. While in the context of mechanical ventilation the term "weaning" comes with an inherent ambiguity due to a number of conflicting definitions [32-36], for the purpose of this activity it was explicitly defined as "the reduction of support delivered by the mechanical ventilator with an end goal of extubation." Consequently, the activity brief prompted participants to annotate periods during which mechanical ventilation weaning takes place and allowed multiple annotations for any individual admission.

Admission Data

The data used in the activity came from a deidentified intensive care dataset called "Medical Information Mart for Intensive Care IV version 2.0" [37]. The eligibility criteria required that subjects were at least 18 years of age at the time of admission, had undergone an invasive mechanical ventilation treatment that lasted for a minimum of 24 hours, their stay in the ICU lasted for a minimum of 4 days, and that their stay did not end with death, including up to 48 hours after discharge. The time-series parameters used in the data extract were limited to those relevant to mechanical ventilation weaning and were selected by 2 independent clinicians working in the ICU. The data were extracted using a structured query language script run on the PostgreSQL installation of Medical Information Mart for Intensive Care IV version 2.0 with the concept tables computed [38]. The script aggregated the selected parameters on an hourly basis for each eligible admission and limited it to the range surrounding the period of mechanical ventilation treatment.

Ethical Considerations

This work was approved by the Faculty of Engineering Research ethics committee at the University of Bristol (case 2022-150, research ethics committee reference 22/HRA/2166). Information regarding the study was circulated with the invited participants electronically alongside the recruitment email. Informed consent was captured electronically using digitally signed consent forms prior to enrollment in the study. Data obtained from this study were deidentified and stored securely in an encrypted database. The access to the database was password-protected and only the research team had access to these data. All participants were able to quit without any explanation at any point during the study. Participation in the study was voluntary, and no financial or other form of compensation was provided to the participants.

Results

Data Annotation Analysis

During the simulated data annotation activity, we observed participants following a series of discrete steps when annotating data. To understand how these actions fit into the annotation process, we analyzed the observations in the context of NIC and aggregated them into the distinct phases of the model.

Planning

Following the distribution of the printout sheets containing the admission data, participants began the process of familiarizing themselves with the data. This stage involved selecting a single excerpt from the available admissions, analyzing the patient demographics, and browsing through the time-series data. During the process, participants frequently cycled through the entire length of the time-series data from the initial admission until discharge and back.

Specifying

Upon familiarization with the specific admission, participants began searching for individual points in time suggestive of the weaning taking place. Some participants preferred to start with the end of treatment, where weaning led to extubation, while others analyzed the data to find the points in time when weaning began. This involved focusing on specific parameters (as suggested by following the specific rows with the pen tip) while browsing through the time-series data across different columns.

Performing

Upon finding the thresholds for the start or end of weaning, the participants annotated the printed sheets directly. To create the labels, a variety of techniques were used, which differed both between the individual annotators and the distinct admissions. These included circling the start and end dates in the header row of corresponding columns, drawing vertical lines on the edges of the columns to mark the start and end periods, and drawing a box around the portion of the dataset spanning the duration of the label. Similarly to the analysis process, some participants preferred to first annotate the end of the weaning process, rather than its start. In some cases, participants also provided additional information in the form of written comments that justified the created label (eg, "mode of ventilation changed from A to B" or "delivered oxygen reduced—indicative of weaning") and underlined values within cells of parameters that prompted annotation.

Perceiving

Following the act of annotating the data, participants frequently verified their labels. This was expressed by browsing through the time-series data again and ensuring that the start and end dates corresponded to the parameter readings that suggested the annotation in the first place.

Interpreting

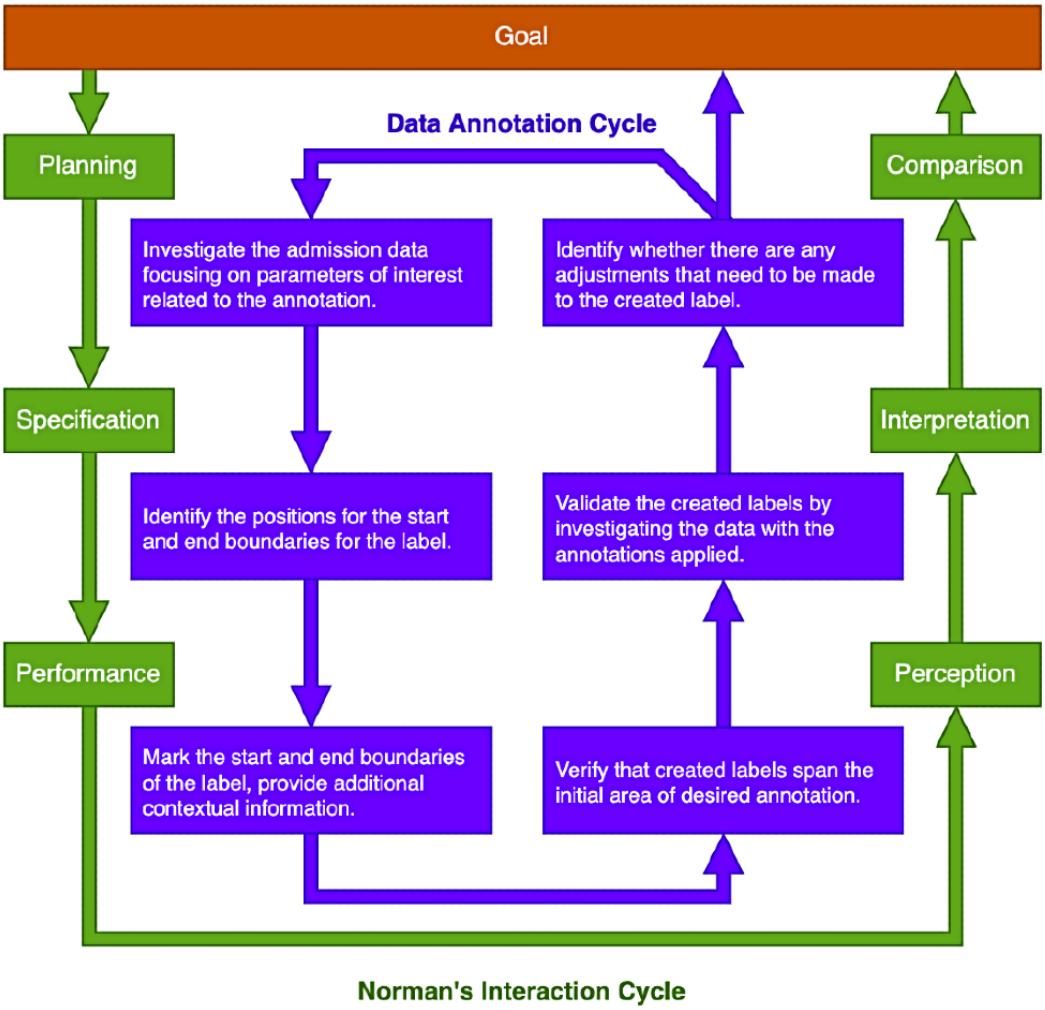
In some cases, participants derived additional insights from the data when investigating them with their labels already present. These insights frequently manifested in the form of retracing the admission data over the span of the created label, which resulted in a reevaluation of the created label and its underlying data.

Comparing

Finally, participants who identified the need for amendments returned to the planning stage to repeat the annotation process, whereas those satisfied with their annotations frequently sought feedback from the activity facilitator on the desired annotation count and the next steps involved in the process.

Crucially, by contextualizing the annotation process within the NIC model we identified that the overall approach to annotation followed a cyclic process of investigation of the data, creation of the labels, reevaluation of the data with the labels applied, and refinement of previously created labels (Figure 4).

Figure 4. Data annotation followed a cyclic process involving investigation of data, label creation, reevaluation of data with labels applied, and refinement of created annotations.



User Requirements

Observations from the workshops captured key characteristics of participant interactions with the task, forming a basis for establishing the functional requirements of the digital tool. Coupled with the operational constraints of data annotation in clinical settings and the specific needs for future label use in the ML context, these characteristics informed the user requirements for the digital annotation tool.

Annotation of Individual Admissions

Familiarization with admission data during the planning and specification stages suggested that the primary goal of participants at the early stage of the process was to obtain a high-level overview of the patient's characteristics and their stay in the ICU. This was supported by the frequent cycling through different parts of the admission, allowing participants to establish a broad timeline of events and the overall treatment trajectory without focusing on the minutia of its delivery. The interface of the tool should allow end users to conduct a similar analysis digitally, incorporating both the patient demographics and the time-series data.

Once familiarized with the admission data, participants shifted their focus to a more task-oriented analysis by narrowing the list of parameters to a subset significant for the task, suggesting

different relevance of specific parameters to the target label. Furthermore, the fact that in certain cases participants found it easier to first search for the end of weaning, rather than its start, suggests that their confidence in the accuracy of the created label, its precise start or end, might be different across annotations. Similarly, each of the participants displayed different preferences for creating annotations favoring first marking either the start of the weaning process or its end. Despite this, they all shared a common set of actions of marking the beginning and end of the label, even when the order of these operations was not the same. To make the annotation process as unobtrusive as possible, the digital data annotation tool should therefore allow flexibility in not only how the data are viewed and analyzed but also how the annotations are being created. As such, the interface needs to display a continuous dataset without constraining it to a subset of time or parameters and allow the end users to freely mark the boundaries of the label in the order of their preference.

Following the creation of a label, participants frequently continued to investigate the underlying data and further adjusted the labels, indicating the need to reflect on their immediate performance and refine the annotations. Adjusting labels upon their reevaluation indicates that when the data are viewed with the annotations applied, the annotator's perception of these data changes allowing for additional insights to be derived. As such, the digital tool should allow for the display of the created annotations on top of the data and facilitate a similar ability to adjust and modify them once they are created.

Participants also frequently provided additional information to support their labels, such as comments or specific parameters that led them to create the annotation. This contextual information could be used as a surrogate for the thought process of the annotators and further inform the importance of specific parameters on the target label. The digital tool should therefore incorporate the ability to provide additional data, such as parameters of interest, confidence in the accuracy of created labels, and free-form text that provides further context.

Finally, the feedback sought upon completing the annotation indicated the need for both a progress indicator and the importance of preservation of the continuity of the annotation process on retaining participants' focus. A digital tool could account for this by displaying the count and information about the already created labels and making it easy and efficient for participants to annotate consecutive admissions. As such, the functional requirements for the digital data annotation tool can be specified as follows:

- R1: Data analysis in which participants are free to navigate through the entire span of the admission data, as well as underlying patient demographics.
- R2: Label creation functionality which is flexible enough to allow end users to select the start and end of annotation independently and in any order.
- R3: Label adjustment capability that facilitates an easy and convenient way to amend the created label, particularly with that label presented on top of data.
- R4: Label supplementation feature that allows participants to provide additional context for the created annotation, including confidence in the label accuracy and relevance of different parameters.
- R5: Workflow continuity that ensures a cohesive process of annotation and keeps the end users engaged and focused on the task.

Semiautomated Annotation

Facilitating a semiautomated approach to annotation that allows for creating labels for an entire dataset is a separate but closely related task. As such, it can benefit from the analysis of the manual annotation process and use it to further inform the requirements for the digital tool. The annotation of a single admission could be described as a bottom-up approach, in which end users are presented with data specific to a single admission and asked to annotate them directly (optimally until the entire dataset is annotated). Conversely, adopting an approach that automates the process would likely involve the annotation of an entire dataset based on a single user input, constituting a top-down approach. A semiautomated approach would therefore focus on creating labels for the whole dataset without focusing on individual admissions during their creation but potentially using them to evaluate the annotations and refine them upon further analysis. To that extent, the digital tool needs to allow its end users to establish annotation strategies that are independent of individual admissions during their formulation.

The process in which participants interacted with the individual admissions suggested several key requirements that needed to be adapted for the semiautomated approach. To close the loop of investigating, annotating, evaluating, and refining that the participants exhibited during the simulated activity, the semiautomated annotation approach facilitated by the platform should also allow for analysis of the effectiveness of the annotation in the context of both the entire dataset and individual admissions. In particular, the interface should allow the end users to view individual admissions with the created annotations applied, providing additional explanations for why the label was created, and in turn, allowing them to refine their annotation strategies. Consequently, the requirements for the digital tool facilitating a semiautomated annotation are defined as follows:

- R6: Annotation and evaluation loop captured as one continuous process that preserves the focus of the end users and facilitates effective annotation.
- R7: Label analysis of individual admissions with the overlay of created annotations on the time-series data and justification for their presence.
- R8: Dataset-wide performance metrics that capture the effectiveness of the annotation in the form of aggregate statistics computed across the entire dataset.

Operational Constraints

In addition to the requirements identified through the analysis of participants' approach to the annotation task, there are further operational constraints that need to be reconciled to deliver an effective annotation workflow. To accommodate the busy schedules of the ICU staff, the tool should allow its end users to access it in a way that suits their needs and does not impose strict time commitments. As such, end users should be able to perform the annotation asynchronously and access the platform

remotely. Furthermore, to preserve the focus and ensure that the end users' time is used efficiently, the system should provide a responsive and time-efficient workflow that enables them to create and evaluate annotations in a short span of time. This is particularly important for the semiautomated approach in which the timely annotation of a large volume of data is critical to preserve the continuity of the annotation and evaluation loop. The operational requirements for the annotation tool are therefore defined as follows:

- R9: Asynchronous and remote annotation that allows the end users to perform the task at their convenience without further complicating their busy schedules.
- R10: Responsiveness of the system that provides feedback on the created annotations in a timely manner, particularly in the case of the semiautomated annotation.

Labels for ML

Finally, having identified the functional and operational requirements, we analyzed the platform in the wider context of using the created labels in ML workflows. To that extent, the discussion between data science experts within our research team surfaced several key factors that the tool should account for to ensure that the captured data can seamlessly integrate into ML pipelines.

The most important one to consider is the need to achieve a careful balance between the number of annotated admissions and the confidence in the created annotations. On one hand, increasing the total number of annotated admissions and their diversity would result in a larger training dataset and, consequently, improve the performance of the resulting ML model. On the other hand, to account for the biases of individual annotators, there needs to be some overlap of admissions annotated by distinct annotators that would allow for a comparison of their approaches and biases. Furthermore, for this to happen, additional metadata surrounding the confidence in the specific annotations and relating the created annotations to their authors also needs to be collected. Only then, will the data scientists be able to evaluate each of the annotators individually, compute their biases, and adjust their annotations accordingly to improve the robustness of the training dataset.

Depending on the volume of the data that needs to be annotated, the nature of the annotation task, and the number and experience of the annotators, this balance might be substantially different. Because of this, the system needs to facilitate a dynamic and adjustable configuration for how specific data are assigned to the annotators. This suggests that in addition to the previously established requirements, the tool should also aim to satisfy the following ML-specific requirement:

- R11: Flexible data-splitting solution that allows for adjustment of the data assigned to each participant and keeps track of the label authorship.

Discussion

Principal Results

This study aimed to explore how intensive care experts approach data annotation to establish requirements for a digital tool designed for annotating large datasets in intensive care. We conducted a manual data annotation activity, observing participants as they completed the task and analyzing their actions to gain insights into their annotation strategies. This analysis informed the development of software requirements for a digital data annotation platform.

Our findings revealed that the time-series annotation process follows a cyclical annotation-evaluation loop, which includes data investigation, label creation, reevaluation of data with the labels applied, and refinement of the created labels. From this analysis, we constructed 11 key requirements: 5 directly related to facilitating the annotation of individual patient admissions, 3 adapted for implementing a semiautomated annotation feature, 2 operational requirements focused on the effectiveness of the annotation workflow, and 1 requirement addressing the future use of labels in the ML context (Table 1).

Table 1. Requirements generated from analysis of observations made during simulated annotation activity.

Category	Requirement description
Annotation of individual admissions	
R1	Data analysis in which participants are free to navigate through the entire span of the admission data, as well as underlying patient demographics.
R2	Label creation functionality which is flexible enough to allow end users to select the start and end of annotation independently and in any order.
R3	Label adjustment capability that facilitates an easy and convenient way to amend the created label, particularly with that label presented on top of data.
R4	Label supplementation feature that allows participants to provide additional context for the created annotation, including confidence in the label accuracy and relevance of different parameters.
R5	Workflow continuity that ensures a cohesive process of annotation and keeps the end users engaged and focused on the task.
Semiautomated annotation	
R6	Annotation and evaluation loop captured as one continuous process that preserves the focus of the end users and facilitates effective annotation.
R7	Label analysis of individual admissions with the overlay of created annotations on the time-series data and justification for their presence.
R8	Dataset-wide performance metrics that capture the effectiveness of the annotation in the form of aggregate statistics computed across the entire dataset.
Operational constraints	
R9	Asynchronous and remote annotation that allows the end users to perform the task at their convenience without further complicating their busy schedules.
R10	Responsiveness of the system that provides feedback on the created annotations in a timely manner, particularly in the case of the semiautomated annotation.
Use in machine learning	
R11	Flexible data-splitting solution that allows for adjustment of the data assigned to each participant and keeps track of the label authorship.

Approach to Annotation

The observations made during the manual annotation activity highlighted important characteristics of how experts from a clinical background approach the annotation task. While the sample of our study produced no discernable differences in how participants from different roles approached the task of data annotation, we observed a variety of strategies used to complete the task. These included several techniques for initial familiarization with the data, prioritization of different parameters during the investigation process, and different order of operations when creating the label. Crucially, we observed that irrespective of the specific strategies, annotators followed a cyclic process of investigation, annotation, reevaluation, and refinement when annotating data. By modeling the process using NIC, we were able to analyze the observations made during the task completion stage and formulate the requirements for the software interface of a digital data annotation platform.

In our analysis, we found that the nature of how annotators choose to analyze the data is largely preferential and, therefore, requires a degree of flexibility in how the software presents the data and allows its users to navigate through it. The cyclic nature of the process suggested that to ensure effective annotation, it must facilitate a continuous and intuitive loop of annotation, allow users to perform actions in varying order, and seamlessly adapt to different stages of the process. Since these requirements were derived from the simulated annotation of individual admissions but captured the characteristics of the overall problem of data annotation, the requirements for the semiautomated approach were adapted to ensure the continuity of the annotation-evaluation loop and flexibility in analysis.

Finally, the fact that annotators frequently sought feedback after creating labels further emphasized the importance of ensuring a seamless and uninterrupted annotation workflow and suggested the need to integrate progress indicators within the digital platform. Together with individual preferences for focusing on different parameters during the analysis and supplying additional information as part of the created labels, this also highlighted the need to capture contextual information alongside labels, which could be beneficial for their later use in the ML pipelines.

Design Implications

This literature on involving clinical staff in the collaborative design of software interventions shows that it can have profound effects on user acceptance and uptake upon release [23,39]. However, it is important to highlight that effectively engaging in such collaboration can be a challenging task, particularly when research activities require substantial time commitment [40]. Collaborating with intensive care clinicians, whose time is in high demand and short supply, may therefore require additional accommodations to be effective. In the context of annotating data, these accommodations translate into additional

requirements that adapt the process to the busy schedules of the clinical staff by allowing them to annotate data both asynchronously and remotely. This further reinforces the need to ensure that the software delivers an intuitive and time-efficient annotation workflow and places additional constraints on its design. Specifically, to facilitate the responsiveness of the annotation-evaluation loop in the semiautomated approach, where users can frequently simultaneously annotate the entire dataset, the software should incorporate a solution that dynamically adapts to both the number of active users and the volume of data that needs to be processed.

Limitations and Future Work

The activity focused on capturing the approach to data annotation was conducted with a limited number of participants, which resulted in a limited diversity of clinical roles and, potentially, perspectives from the ICU. This suggests that some of the requirements established in this research could be particularly applicable to annotators in junior doctor positions and therefore not necessarily generalizable across entire staff working in ICUs. The selection of the task for the manual annotation activity was made to provide an annotation experience that could be extrapolated beyond the specifics of the task. Despite this, we acknowledge that the results it produced could be biased specifically toward the annotation of weaning from mechanical ventilation.

Furthermore, our approach to capturing the annotation process during the simulated activity with only a single observer had a significant impact on the quantity and quality of observations that were collected. Due to the imbalance in the number of annotators and observers, some of the actions undertaken by the participants during the annotation could have gone unnoticed. Furthermore, the presence of a single observer created an inherent bias in the collected observations, as the observed actions could have been perceived differently by observers with different backgrounds or characteristics. These limitations suggest that the observations collected during the activity come with a degree of incompleteness and inaccuracy, which may have impacted the elicited requirements.

We acknowledge that further work in establishing the requirements for the digital tool could strengthen the understanding of how clinicians approach the data annotation task. To that extent, we suggest that further research in this area focuses on capturing the requirements that expand beyond the confines of a single annotation task and within a broader and more diverse population from the ICUs. Capturing a wider range of perspectives could inform the applicability of the elicited requirements and, in consequence, strengthen the resulting design of the digital tool.

The nature of the activity itself was focused strictly on the direct annotation of individual admissions rather than the use of any assistive or automated technologies. Although several requirements elicited in this context applied to the overall annotation process, including a semiautomated approach, additional requirements not captured in this study may also exist. Further research should therefore focus on the evaluation of the proposed requirements and their use and limitations in real-world applications. Therefore, these requirements should be used to design and implement a digital data annotation platform that should be trialed within a clinical setting. Conducting a study that investigates the feasibility of a semiautomated approach to the data annotation, particularly in comparison with the direct annotation of individual admissions, could further inform the requirements for data annotation tools.

Conclusions

In this study, we investigated how clinical staff from ICUs approach the task of data annotation and established 11 key requirements for a digital data annotation tool that could be deployed within the health care setting. Our findings revealed that data annotation is a cyclic process that demands flexibility in how annotators investigate and annotate the data. Preservation of the workflow continuity across different admissions and fluid transition between analysis, annotation, and refinement of the label are essential to facilitating effective data annotation in the clinical domain. Adaptations for the semiautomated annotation need to consider these factors by providing a responsive interface that dynamically adapts to the volume of data and allows for analysis on both an individual and a dataset-wide basis. The significance of these findings is evident in their potential to guide the development of a data annotation tool that capitalizes on the considerable data generated in the ICUs and the expanding use of computational methods in health care.

Acknowledgments

This work was partly supported by the Engineering and Physical Sciences Research Council Digital Health and Care Centre for Doctoral Training at the University of Bristol (UK Research and Innovation grant EP/S023704/1 to MW). The work of RSR was funded by the UK Research and Innovation Turing AI Fellowship (grant EP/V024817/1). The project received an AWS Cloud Credit for Research grant.

Authors' Contributions

MW was involved in the study design, workshop facilitation, and writing. RSR, CMW, and CB were involved in the study design and critical revision.

Conflicts of Interest

None declared.

References

1. Russell SJ, Norvig P. Artificial Intelligence: A Modern Approach, Global Edition. In: Prentice Hall Series in Artificial Intelligence. Singapore Taipei Tokyo: Pearson; 2016.
2. Bellman R. An Introduction to Artificial Intelligence: Can Computers Think?. San Francisco: Boyd & Fraser Publishing Company; 1978.
3. Grosan C, Abraham A. Intelligent Systems: A Modern Approach. In: Intelligent Systems Reference Library. Berlin: Springer-Verlag; 2011.
4. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 1982;79(8):2554-2558 [FREE Full text] [doi: 10.1073/pnas.79.8.2554] [Medline: 6953413]
5. Introduction TML, Fourth E. Introduction to Machine Learning, Fourth Edition. In: Adaptive Computation and Machine Learning Series. Cambridge, MA: MIT Press; 2020.
6. Jain A. Overview and importance of data quality for machine learning tasks. 2020 Presented at: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Virtual Event CA USA: ACM; 10 Jan 2025; United States p. 3561-3562. [doi: 10.1145/3394486.3406477]
7. Noorbakhsh-Sabet N, Zand R, Zhang Y, Abedi V. Artificial intelligence transforms the future of health care. Am J Med 2019;132(7):795-801 [FREE Full text] [doi: 10.1016/j.amjmed.2019.01.017] [Medline: 30710543]
8. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
9. Chapalain X, Huet O. Is artificial intelligence (AI) at the doorstep of intensive care units (ICU) and operating room (OR)? Anaesth Crit Care Pain Med 2019;38(4):337-338. [doi: 10.1016/j.accpm.2019.05.003] [Medline: 31102792]
10. Hunter B, Hindocha S, Lee RW. The role of artificial intelligence in early cancer diagnosis. Cancers (Basel) 2022;14(6):1524 [FREE Full text] [doi: 10.3390/cancers14061524] [Medline: 35326674]
11. Heo MS, Kim J, Hwang J, Han S, Kim J, Yi W, et al. Artificial intelligence in oral and maxillofacial radiology: what is currently possible? Dentomaxillofac Radiol 2021;50(3):20200375 [FREE Full text] [doi: 10.1259/dmfr.20200375] [Medline: 33197209]
12. Azizi S. Big self-supervised models advance medical image classification. 2021 Presented at: 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 10 Oct 2021; Montreal, QC, Canada p. 3458-3468. [doi: 10.1109/iccv48922.2021.00346]
13. Martinez-Martin N, Luo Z, Kaushal A, Adeli E, Haque A, Kelly SS, et al. Ethical issues in using ambient intelligence in health-care settings. Lancet Digital Health 2021;3(2):e115-e123. [doi: 10.1016/s2589-7500(20)30275-2]
14. Burki TK. Artificial intelligence hold promise in the ICU. Lancet Respir Med 2021;9(8):826-828. [doi: 10.1016/s2213-2600(21)00317-9]
15. Gautam S, Srinath M. Blind spots and biases: exploring the role of annotator cognitive biases in NLP. arXiv Preprint posted online on April 29, 2024 [FREE Full text] [doi: 10.18653/v1/2024.hcinlp-1.8]
16. Spasic I, Nenadic G. Clinical text data in machine learning: systematic review. JMIR Med Inform 2020;8(3):e17984 [FREE Full text] [doi: 10.2196/17984] [Medline: 32229465]
17. Chen Y, Lasko TA, Mei Q, Denny JC, Xu H. A study of active learning methods for named entity recognition in clinical text. J Biomed Inform 2015;58:11-18 [FREE Full text] [doi: 10.1016/j.jbi.2015.09.010] [Medline: 26385377]
18. Kholghi M, Sitbon L, Zuccon G, Nguyen A. Active learning: a step towards automating medical concept extraction. J Am Med Inform Assoc 2016;23(2):289-296 [FREE Full text] [doi: 10.1093/jamia/ocv069] [Medline: 26253132]
19. South BR, Mowery D, Suo Y, Leng J, Ferrández Ó, Meystre SM, et al. Evaluating the effects of machine pre-annotation and an interactive annotation interface on manual de-identification of clinical text. J Biomed Inform 2014;50:162-172 [FREE Full text] [doi: 10.1016/j.jbi.2014.05.002] [Medline: 24859155]
20. Pezoulas VC, Zaridis DI, Mylona E, Androutsos C, Apostolidis K, Tachos NS, et al. Synthetic data generation methods in healthcare: a review on open-source tools and methods. Comput Struct Biotechnol J 2024;23:2892-2910 [FREE Full text] [doi: 10.1016/j.csbj.2024.07.005] [Medline: 39108677]
21. Vandekerckhove P, de Mul M, Bramer WM, de Bont AA. Generative participatory design methodology to develop electronic health interventions: systematic literature review. J Med Internet Res 2020;22(4):e13780 [FREE Full text] [doi: 10.2196/13780] [Medline: 32338617]
22. Kautz K. Participatory design activities and agile software development in human benefit through the diffusion of information systems design science research. In: Pries-Heje J, Venable J, Bunker D, Russo NL, DeGross JI, editors. IFIP Advances in Information and Communication Technology. Berlin, Heidelberg: Springer; 2020:303-316.
23. Damodaran L. User involvement in the systems design process-a practical guide for users. Behav Inf Technol 1996;15(6):363-377. [doi: 10.1080/014492996120049]
24. Sanz MF, Acha BV, García MF. Co-design for people-centred care digital solutions: a literature review. Int J Integr Care 2021;21(2):16 [FREE Full text] [doi: 10.5334/ijic.5573] [Medline: 33981193]
25. Curtis K, Brooks S. Digital health technology: factors affecting implementation in nursing homes. Nurs Older People 2020;32(2):14-21. [doi: 10.7748/nop.2020.e1236] [Medline: 32159302]

26. Hobson EV, Baird WO, Partridge R, Cooper CL, Mawson S, Quinn A, et al. The TiM system: developing a novel telehealth service to improve access to specialist care in motor neurone disease using user-centered design. Amyotroph Lateral Scler Frontotemporal Degener 2018;19(5-6):351-361 [FREE Full text] [doi: 10.1080/21678421.2018.1440408] [Medline: 29451026]
27. Norman DA. The Design of Everyday Things, Revised and Expandes Editons. Cambridge, MA London: MIT Press; 2013.
28. Soegaard M, Dam RF. Gulf of evaluation and gulf of execution. Interaction Design Foundation. 2015. URL: https://www.interaction-design.org/literature/book/the-glossary-of-human-computer-interaction/gulf-of-evaluation-and-gulf-of-execution [accessed 2025-09-10]
29. Slutsky AS, Hudson LD. Mechanical ventilation. In: Goldman's Cecil Medicine. Philadelphia: Elsevier; 2012:638-645.
30. Girard TD. Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (awakening and breathing controlled trial): a randomised controlled trial. Lancet (London, England) 2008;371(9607):126-134. [doi: 10.3410/f.1097928.562141]
31. Nichol AD, Geoghegan P, Duff S, Cooper DJ, Devlin D. What is the optimal approach to weaning and liberation from mechanical ventilation? Evidence-Based Practice of Critical Care.: Elsevier; 2019. URL: https://research.monash.edu/en/publications/what-is-the-optimal-approach-to-weaning-and-liberation-from-mecha-2 [accessed 2025-09-10]
32. McConville JF, Kress JP. Weaning patients from the ventilator. N Engl J Med 2012;367(23):2233-2239. [doi: 10.1056/nejmra1203367]
33. Nitta K, Okamoto K, Imamura H, Mochizuki K, Takayama H, Kamijo H, et al. A comprehensive protocol for ventilator weaning and extubation: a prospective observational study. J Intensive Care 2019;7(1):50 [FREE Full text] [doi: 10.1186/s40560-019-0402-4] [Medline: 31719990]
34. Mador MJ. Weaning from mechanical ventilation: what have we learned and what do we still need to know? Chest 1998;114(3):672-674. [doi: 10.1378/chest.114.3.672] [Medline: 9743148]
35. Boles JM, Bion J, Connors A, Herridge M, Marsh B, Melot C, et al. Weaning from mechanical ventilation. Eur Respir J 2007;29(5):1033-1056 [FREE Full text] [doi: 10.1183/09031936.00010206] [Medline: 17470624]
36. Lermitte J, Garfield MJ. Weaning from mechanical ventilation. Contin Educ Anaesth Crit Care Pain 2005;5(4):113-117. [doi: 10.1093/bjaceaccp/mki031]
37. Johnson A, Bulgarelli L, Pollard T, Gow B, Moody B, Horng S, et al. MIMIC-IV. PhysioNet. 2022. URL: https://physionet.org/content/mimiciv/3.1/ [accessed 2025-01-10]
38. Johnson AEW, Stone DJ, Celi LA, Pollard TJ. The MIMIC code repository: enabling reproducibility in critical care research. J Am Med Inform Assoc 2018;25(1):32-39 [FREE Full text] [doi: 10.1093/jamia/ocx084] [Medline: 29036464]
39. Slattery P, Saeri AK, Bragge P. Research co-design in health: a rapid overview of reviews. Health Res Policy Syst 2020;18(1):17 [FREE Full text] [doi: 10.1186/s12961-020-0528-9] [Medline: 32046728]
40. Kushniruk A, Hr C. Participatory design, user involvement and health IT evaluation, in evidence-based health informatics. In: Studies in Health Technology and Informatics. Amsterdam: IOS Press; 2016:139-151.

Abbreviations

AI: artificial intelligence
ICU: intensive care unit
ML: machine learning
NIC: Norman's Interaction Cycle

Edited by A Choudhury; submitted 29.01.24; peer-reviewed by T Abd El-Hafeez, S Tayebi Arasteh, M Lotfinia; comments to author 11.10.24; revised version received 05.12.24; accepted 01.01.25; published 05.02.25.

Please cite as:
Wac M, Santos-Rodriguez R, McWilliams C, Bourdeaux C
Capturing Requirements for a Data Annotation Tool for Intensive Care: Experimental User-Centered Design Study
JMIR Hum Factors 2025;12:e56880
URL: https://humanfactors.jmir.org/2025/1/e56880
doi:10.2196/56880
PMID:

User-Centered Prototype Design of a Health Care Robot for Treating Type 2 Diabetes in the Community Pharmacy: Development and Usability Study

Ching-Ju Chiu[1], PhD; Lin-Chun Hua[1], MS; Jung-Hsien Chiang[2], PhD; Chieh-Ying Chou[1,3], MD, MS

[1]Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

[2]Department of Computer Science and Information Engineering, National Cheng Kung University, Tainan City, Taiwan

[3]Department of Family Medicine, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, No. 138, Sheng Li Road, Tainan City, Taiwan

Corresponding Author:
Chieh-Ying Chou, MD, MS
Institute of Gerontology, College of Medicine, National Cheng Kung University, Tainan City, Taiwan

Abstract

Background: Technology can be an effective tool for providing health services and disease self-management, especially in diabetes care. Technology tools for disease self-management include health-related applications for computers and smartphones as well as the use of robots. To provide a more effective continuity of care and to better understand and facilitate disease management in middle-aged and older adult patients with diabetes, robots can be used to improve the quality of care and supplement community health resources, such as community pharmacies.

Objective: The aim of this study was to develop a health care robot prototype that can be integrated into current community pharmacies.

Methods: Three user-centered approaches were used: (1) review of the literature on technology use among older adults, 2) reference to the seven key diabetes self-care behaviors by the American Association of Diabetes Educators (AADE), and (3) meeting with health care providers in the community. Field investigations and interviews were conducted at community pharmacies and diabetes health education centers to determine the appearance, interface, content, and function of the robot.

Results: The results show that diabetes health care prototype robots can be established through user-centered design. The following important features were revealed: (1) perceived ease of use is considered a friendly operating interface; therefore, we used less than 3 buttons in an interface; (2) minimization of the interface between blue and yellow, which is unfriendly to older adults; (3) the health education mode was the most preferred mode with sound, image, and video presentation; (4) the most predilected functions are health education resources and health records, and that patient data can be easily collected through health education games and dialogue with robots; and (5) touching the screen is the most preferred operation mode.

Conclusions: An evidence-based health care robot can be developed through user-centered design, an approach in which a model that connects medical needs to people with health conditions can be built, thereby facilitating the sustainable development of technology in the diabetes care field.

(JMIR Hum Factors 2025;12:e48226) doi:10.2196/48226

KEYWORDS

robot; diabetes; self-management; middle-aged adult; community pharmacy; older adult; prototype

Introduction

Diabetes is a significant global health issue, with the number of people affected by the disease continually rising [1]. According to the International Diabetes Federation Diabetes Atlas, the global number of people with diabetes reached 537 million in 2021 and is expected to continue rising, reaching an estimated 783 million by 2045 [2]. The most prevalent form of diabetes is type 2 diabetes, which affects more than 90% of all individuals with diabetes [2]. Uncontrolled diabetes over time can lead to severe complications, including cardiovascular diseases, lower-limb amputation, kidney damage (nephropathy), nerve damage (neuropathy), and blindness (retinopathy) [2,3]. Age is one of the risk factors for type 2 diabetes, with its prevalence and mortality rates rising as individuals get older [4,5]. According to statistics from Taiwan's Ministry of Health and Welfare, the prevalence of diabetes increases significantly from the age of 45 years for both men and women [6]. As the duration of diabetes and patient age increase, patients with diabetes face not only a higher incidence of complications but also challenges related to multiple medications and comorbidities [7]. This underscores the importance of focused

health care management for middle-aged and older adult patients with diabetes.

In recent years, technology has emerged as a crucial tool for disease management, including diabetes [8,9]. The demand for digital health care has surged, evidenced by the introduction of more than 90,000 new health-related apps on platforms like the Apple Store and Google Play in 2020 alone, contributing to a total of over 350,000 health care applications available globally [10]. Among these technological advancements in health applications, health care robots are one of the future trends in health technology [11], driven by the ongoing digitization of health care services.

Community pharmacies play a vital role in providing health care services, offering convenient access to medications and health services that support local health care needs. There are currently more than 7400 health insurance pharmacies across Taiwan [12], and a 2016 survey indicated that over 80% of these pharmacies are open for at least 12 hours, with around 40% operating almost year-round [13]. Compared to hospitals, community pharmacies face less pressure from time constraints and space limitations, positioning them as key players in bridging the gap between individuals and larger medical institutions.

Despite the availability of professional medical information in clinical settings, there remains a challenge in ensuring that patients effectively absorb and apply health education and professional advice once they return to their communities. This gap highlights the need for technological tools that support continuous care, especially for middle-aged and older adults managing diabetes. Integrating robots into community health resources, such as community pharmacies, could enhance the quality of care, improve health management, and advance health promotion within the community. The expertise of pharmacists, coupled with the high accessibility of community pharmacies, makes them an ideal setting for such interventions.

However, current technology development for disease management often excludes medical professionals from the design process and lacks real-world testing [14]. Additionally, there is no clear evidence that technology can effectively integrate with community medical resources or bridge the gap between medical institutions and personal care. Therefore, the aim of this study was to design and develop a prototype of a diabetes care robot that meets the needs of middle-aged and older adult patients with diabetes, aligns with the perspective of community pharmacist care, and can be integrated into the current community pharmacy setting.

Methods

Participants

The design and development of the prototype robot in this study were primarily conducted through literature-based theories and expert team meetings [15]. The development of the prototype robot involved the researcher synthesizing the literature and conducting field visits to pharmacies and health education units. The development goals and direction were then jointly decided during team meetings. The development of the prototype robot began in March 2018. The team included 1 expert in technology, 1 expert in gerontology, 2 physicians, 12 to 15 engineers and research assistants, and 1 researcher. To maintain progress and continuously refine the prototype robot, the team held regular meetings 1 - 2 times per month until a more complete interactive functionality was established to enhance the user experience. Furthermore, to gain a more realistic understanding of the feasibility and initial results of the robot's application in community health care, 30 middle-aged and older adult patients with diabetes, along with 10 community pharmacists, were also included to provide user feedback.

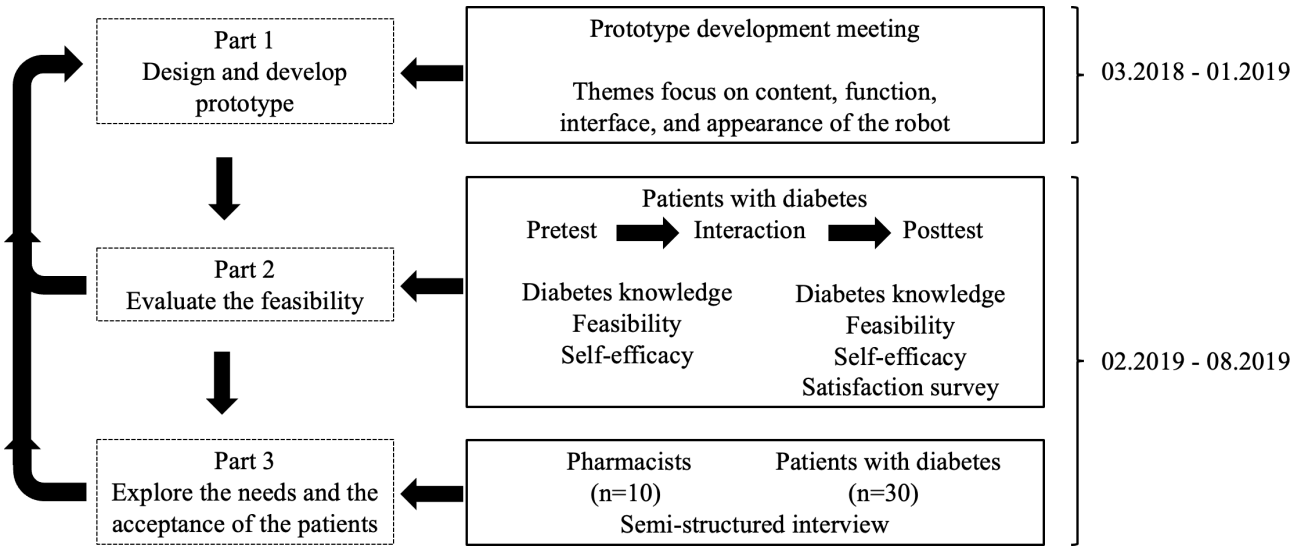
Ethical Considerations

This study was performed in accordance with the relevant guidelines and regulations, including the Declaration of Helsinki and was approved by the Institution Review Board (IRB) of National Cheng Kung University Hospital in Taiwan (No. A-ER-105 - 509). All the participants provided signed informed consent, and all study data were deidentified to protect the identity of participants. Participants did not receive any financial or material compensation for their participation in this study.

Procedures

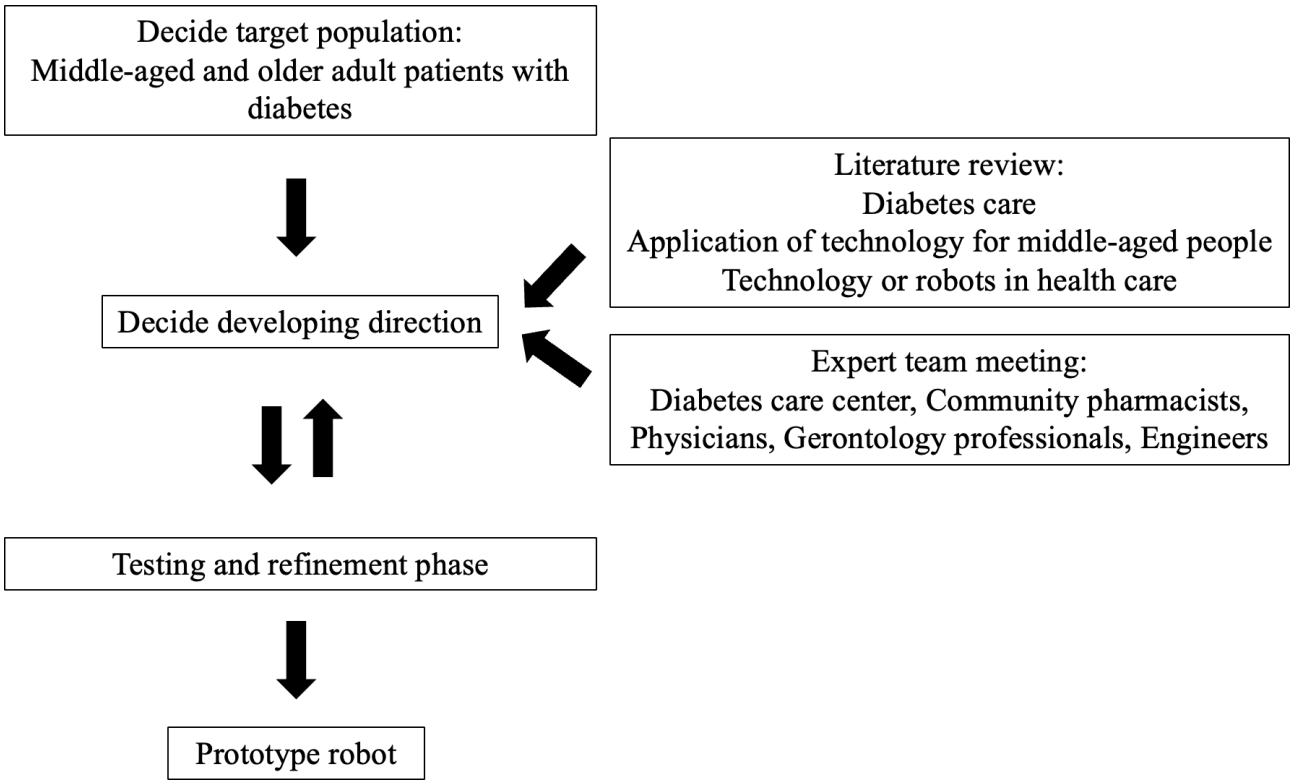
The research design framework for the prototype robot comprised three steps: using an evidence-based literature review and team meeting to design and develop the robot prototype, evaluating feasibility through collecting pre- and postassessment questionnaires, and exploring needs and acceptance through interviews with the target population, as illustrated in Figure 1.

Figure 1. The research design framework for the prototype robot used in community-based diabetes care.



Although robotics is an emerging technology, there is a lack of understanding about the applications of robotics. To gain a better understanding of user needs, a high-fidelity prototype can be used to provide users with real-world experience and stimulate their needs, thus enabling them to provide more in-depth insights for future development and design [16]. Therefore, the first part of the study was to design and build a prototype community-based diabetes care robot to facilitate user experience, as illustrated in the detailed flowchart shown in Figure 2.

Figure 2. Process showing how to build a prototype community-based diabetes care robot.



The second and third parts were dedicated to testing and evaluating the prototype robot. In the second part, it took 15 - 20 min for the patients with diabetes to interact with the robot. Before and after the interaction, questionnaires including a diabetes knowledge test, self-efficacy for diabetes, and feasibility of use of the robot were administered. In the third section, semistructured qualitative interviews were conducted with middle-aged and older adult patients with diabetes and community pharmacists about their experiences and visions for using the robots. The detailed methods and results of the pre- and postassessment questionnaires, including a diabetes knowledge test, self-efficacy for diabetes, and the feasibility of using the robot in patients with diabetes, as well as the in-depth interviews with both pharmacists and patients with diabetes, were addressed in our previously published study [17]. This prior study demonstrated that robot interaction significantly improved patients' diabetes knowledge and the perceived feasibility of its use in pharmacies. Both patients and

pharmacists gave positive feedback, with patients valuing self-directed learning, comfortable interaction, and vivid engagement, while pharmacists highlighted real-world applicability and new service potential. These findings offer foundational insights that inform the design and objectives of our current research.

Design and Construction of a Prototype Community-Based Diabetes Care Robot

As the goal of this research is to develop a prototype for a community-based diabetes care robot and to gather and understand the perspectives and needs of patients with diabetes and community health care providers, the initial team meetings decided to use an existing robot with foundational functions. This robot was optimized and gradually modified to develop interactive features relevant to diabetes care, aiming for efficiency and smooth user operation. The robot developed in this study uses Zenbo (ASUS), a smart home product from ASUS. The appearance and specifications are shown in Figure 3 and Textbox 1, respectively.

Figure 3. Appearance of the Zenbo robot.

Textbox 1. Zenbo robot specifications and functions; source: ASUS.

Specifications:

- External dimensions: 37 × 37 × 62 cm (length × width × height)
- Weight: 10 kg
- Operating system: Android
- Display screen: 10.1" LCD screen
- Language skills: Chinese
- Functions: Listening, speaking, seeing, emotional expression, playing music and videos, movement, internet connection, and learning

LCD: liquid crystal display

The concept for the prototype robot is primarily derived from research and theory, which must align with the needs of the target population, namely middle-aged and older adult patients with diabetes. Therefore, this study focused on theories and care indicators related to diabetes care [18], collected and organized the design and limitations of technology applications for middle-aged individuals [19,20], and examined the development and application of current technology or robots in health care, as well as the connection between the use of technology and health in middle-aged people and the mode for delivering health services [21], to establish the development objectives [22,23]. Three user-centered approaches were used in the design and construction of a prototype community-based diabetes care robot: (1) a review of the literature on technology use among older adults, (2) reference to the seven key diabetes self-care behaviors by the American Association of Diabetes Educators (AADE), and (3) meetings with health care providers in the community. The literature review search strategy of this study involved collecting studies related to the application of robots in the elderly population [24-27]. Field investigations and interviews were conducted at community pharmacies and diabetes health education centers to determine the design of the robot. Because patients with diabetes go to community pharmacies to take their medications, and in addition to providing pharmaceutical services, community pharmacies can also offer related information on disease care. Therefore, in addition to collecting service experiences from the community pharmacists, the choice of a community pharmacy as the testing site for the prototype is driven by the fact that these pharmacies provide a secure, private, and professionally consultative environment.

The design framework of the community-based diabetes care robot was devised to collect and compile data in four main areas: appearance, interface, content, and function. For appearance, we collected data about the appearance and size of the robot. For interface, data were collected about the color, text size, arrangement, and other screen design. For content, the system was responsible for building the information that needed to be conveyed, with the user's needs as the primary consideration. For function, we collected data about the presentation of the interactive process and information transfer.

Once the direction of development has been established, a prototype robot is initially built that can be used in practice. The development of technology does not only rely on past and proven experience, but also must be tailored to the needs of real users. Therefore, the prototypes need to interact with the users in real life and give them a real experience in accordance with the proposed development directions. It is only after this that users can give practical advice.

Results

Using a literature search and practical experience, this study focused on data collection in four main areas: appearance, interface, content, and function of robot design. The development directions and options for prototype robots are listed in Table 1.

Table . Development directions and options for prototype robots.

	Literature and theory	Interviews and team discussions	Results and design directions
Appearance	• Type: The humanoid type is less well accepted. The doll type, on the other hand, is seen as a toy for children [22]. Prototype robots should avoid being too complex or trendy. However, if the prototype lacks a humanoid feel, it will not be favored by senior citizens [23]. • Application mode: The focus of the robot varies between different appearances. Appearance will affect whether it is presented in a functional or interactive way [28,29].	• Size: The prototype robot must fit into the space of the pharmacy. • Height: It should be at a height at which middle-aged to older adults can sit and see it horizontally.	• It must be confirmed that this model is not complicated and is expected to be well-accepted because it is a non-humanoid model. However, it may be seen as a toy because of the cute non-humanoid appearance. The model does not have hands and feet to operate and therefore operates mainly in a conversational manner. It must be confirmed that the angle and height of the interaction mode (eg, the height of the table or chair on which the robot can be placed) can be adjusted.
Interface	• Colors: Blue and yellow are less recognizable [19]. Yellow and green combinations, and red and purple combinations, are prone to recognition errors [20]. • Operation: Per the Technology and Acceptance Model [30], the interface and operation must be simple because it is used by older adults. The font should be large and clear. The interface should have a maximum of four buttons. All other interfaces have two buttons for simple selection [31].	• The interface should be simple. A single screen does not need much emphasis. Text should be enlarged.	• The color scheme of the interface is highly contrasting to avoid confusion over similar color schemes. The use of blue and yellow colors is reduced and the number of buttons on the interface is 2 - 3. The content is centered, and apart from the main screen, the content is mostly presented as a single theme. The font size ranges from a minimum of 30sp to a maximum of 50sp (sp being the text unit in the Android system).
Content	• Seven key diabetes self-care behaviors by the American Association of Diabetes Educators [18]	• Applications addressing various everyday challenges and health education videos on common diseases in community pharmacies.	• The content of the system is based on the seven diabetes self-care behaviors, with games and interactive dialogues to present a variety of health education methods. To make the content more relevant to the public, common disease symptoms and complications of diabetes are presented.
Function	• Operation: Technology and Acceptance Model [30]. • Health education: Multisensory perception stimulates learning [21], with a core health education focus each time. • Technology applications: The applications of robots in health management can include disease management, monitoring of physiological data, and interactive health education [26,32,33]. The use of technology in diabetes health management focuses on recording and reminding, as well as providing information and entertainment [34-36].	• There are games, friendly greetings, and flexible interaction times.	• The health education mode is presented with sound, images, and video, avoiding a focus on the use of a single sense. The functionality of the current model will be considered and the information on functionality available in the literature will be further researched.

Appearance

Although the robot's appearance was determined based on the available resources, the chosen model features a cute, non-humanoid design to better align with the preferences of older adults and enhance user acceptance. Although the robot's height is a limiting factor, placing the robot on elevated surfaces, such as tables or chairs, can correct height-related issues, making it easier for users to operate.

Interface

As the interface is critical to the overall smoothness of the robot, the presentation of the interactive interface must meet the needs of middle-aged and older adults in terms of the ease of use of technology products and text reading as much as possible. In addition, according to the Technology and Acceptance Model [30], the perceived ease of use should be considered as a whole and should take into account the user-friendliness of the interface so that the user can experience the clearest and easiest process.

In the interface development, to make it easiest for older adults to operate, each screen will display a maximum of two main options, with options clearly framed to reduce the likelihood of users being unsure of their choices. Information and selection buttons will be centrally placed and maximized on the screen to help users quickly focus on key points. The main screen will feature simple text and image displays. To avoid causing difficulties in recognition for older users, the background will be a single color, and the overall interface will use no more than two main color tones. The main screen and option areas will predominantly use a black-on-white color scheme, with high contrast to make text easier to read.

Content

The main content of the development includes the seven key diabetes self-care behaviors according to the AADE standards. The overall main content and functions presented by the prototype robot are detailed in Table 2 and aligned with the seven AADE diabetes care indicators.

The main functions of the system are health education resources and health records. Health education resources are presented through health education quizzes, pamphlets, and videos. The health records section includes health values, discomfort, and nutrition. The concept of this record is to help patients with diabetes manage their records, including basic physiological data such as blood pressure, blood glucose, and blood lipids. The relevant data are entered and the corresponding standard values are provided. The focus of the "Discomfort" program is to understand and record the various conditions in the patients with diabetes and to compare the complications associated with diabetes with the common problems of the family. By reminding and advising patients to be more alert to their health conditions, they can gain a better understanding of their health conditions. The results are recorded for follow-up. All interactions can be fed back to the user in the form of paper or communication software messages.

Table . Functional development and presentation of prototype robots.

Theme	Subtheme	Theme content	Presented at the American Association of Diabetes Educators
Health education resources	Health education test	Questions focus on the seven major areas of diabetes care and common diabetes health problems	Healthy eating, regular exercise, blood glucose monitoring, proper medication use, problem solving, health adaptation, and risk reduction
	Health education leaflet	Share and recommend links to help filter relevant content	Healthy eating, regular exercise, blood glucose monitoring, proper medication use, problem solving, health adaptation, and risk reduction
	Health education videos	Share and recommend links to help filter relevant content	Healthy eating, regular exercise, blood glucose monitoring, proper medication use, problem solving, health adaptation, and risk reduction
Health records	Health values	Alert and record blood pressure, blood glucose and lipids, and provide standard values	Blood glucose monitoring
	Discomfort	Presented in a lifestyle format to remind patients of complications and other problems they may encounter and link them to their disease conditions	Problem solving, health adaptations, and risk reduction
	Nutrition	Ask each other about their dietary status in a simple conversation to help them develop a balanced diet	Healthy eating

Function

To establish the overall function, we should first understand the current mode of application of technology in disease and health management, including the application of health technology for older adults, diabetes management technology tools, and the application of robots in health management, and consider the current operation fluency and feasibility of using robots.

According to the current functional status of the robot, the main functions are primarily simple question-and-answer interactions with pictures and images, so the main functions are presented in a health education manner. A game was used to transfer knowledge on diabetes-related care content, and a thematic health education video was shown as feedback to the patients according to their wrong answers.

The main objective of the development of this prototype robot is to enable patients with diabetes to experience actual interaction with the robot and to understand the possible interaction patterns, in order to understand the experience and perspective of the user. During the development process, the prototype robot is adapted to meet the needs of the community and medical personnel based on the health education needs of the Diabetes Health Centre and the feasible applications of the community pharmacy. In view of the time and cost of development as well as the operation of the technology, the functions were not expanded extensively and were mainly presented in the form of a basic framework. After the patient with diabetes or community health care worker has confirmed the suitability of the directions and content, the functions will be expanded and built. To allow each participant to experience the same features, the most stable and fully developed part (ie, the health education test, a game with the robot, that focuses on diabetes-related knowledge) is used as the main axis for participant interaction. The rest of the developed functions are presented as introductions and simple demonstrations.

Discussion

Principal Findings

This study demonstrated that an evidence-based health care robot can be developed through user-centered design, facilitating the sustainable advancement of technology in the field of diabetes care. The use of technology in patient management can be beneficial to patients. However, the applications of technology that are readily available on the market are not always effective. While the current technology interventions are mostly aimed at behavioral change, they are rarely supported by therapeutic guidelines and theories [35,37], making the development process too subjective or deviating from reality. The core elements of technology development must therefore be user-centered. Thus, based on the technology being developed and designed for diabetes disease self-management, Goyal et al proposed four stages of user-centered development [38]. The first is the so-called development process, which must be evidence-based and analyze possible barriers and problems and must be culturally and environmentally appropriate. In the second stage, the feasibility of the system must be understood. The third stage involves further evaluation of effectiveness. The fourth stage requires more long-term monitoring. At the same time, as the four stages are in a continuous cycle and interact with each other, it is necessary to revise and review them repeatedly to build a technology tool that meets the needs of users [38]. Furthermore, personal health information must also be protected. While this prototype requires login, ongoing efforts should be made to enhance and ensure information security.

Comparison With Prior Work

In previous studies of diabetes care robots, the main focus has been on younger age groups of patients with type 1 diabetes [32,33,39], and less on middle-aged patients with type 2 diabetes. Even in the case of older adult care, the functional applications of robots are mainly companionship-based or focus on social functions for older adults. There should be less involvement in behavioral change, but this is also an important part of health messaging. Therefore, this study uses the seven key diabetes self-care behaviors by the AADE to consider the basics of healthy eating, regular exercise, blood glucose monitoring, and proper medication use as core health-enhancing skills, and to help patients with diabetes better understand their disease and their health. These seven self-care behaviors are interlinked with diabetes disease self-management and not only directly contribute to disease management (eg, reminders and records), but also enhance the overall quality of care for the disease by building up the correct concepts of patients with diabetes. This content is presented as a basic health education leaflet and a health education video to convey information. The health education resources are sourced from relevant health education websites and reference links from major hospitals and institutes, and can be categorized to help patients quickly search and link to relevant functions. In terms of the functions, we should first understand the current mode of application of technology in disease and health management, including the application of health technology for older adults, diabetes management technology tools, and the application of robots in health management, and consider the current operation fluency and feasibility of using robots. The robot itself has basic functions such as playing videos and documents. Based on the current functionality of the robot, the main focus of further development is on the simple interaction with pictures and images. The main function is therefore presented in the form of health education. To make the health education interaction more efficient and interesting, a health education game was used to transfer knowledge about diabetes-related care content, and at the end of one stage of the game, a health education video on the relevant topic was selected and shown as feedback based on the questions that the patient answered incorrectly, to deepen the patient's understanding of the concept.

Past studies have also supported the idea that a multimedia approach to health education can help patients understand the content of health education [40,41]. The health education model is more stimulating, as it delivers information to patients in a variety of ways, such as videos and animations, which in turn produces better health education results. This is one of the main advantages of robots for the implementation of health education [40]. The robot's rich expressions and gestures, as well as its voice-to-voice interactions, are an advantage in teaching [41]. The prototype utilized in this study is a nonhumanoid robot. While humanoid robot interactions are more human-like, their realistic appearance can often induce fear, leading to lower acceptance. In contrast, nonhumanoid robots, though cute, might be perceived more like toys, resembling a game. The advantage lies in a more engaging interaction, but the downside is that they may be seen merely as playthings, potentially overshadowing the educational information they convey.

In the delivery of health services by robots, there is a need for a thorough consideration of the technological side, the professional side, and the patient side. Therefore, in addition to the professional development in the field of technology, the initial development should also identify and create a suitable development direction from the perspective of medical professionals, and build a prototype robot. Later on, the prototype robot can be used to enable patients with diabetes to experience the interaction with the robot, and through this interaction, to gather more in-depth information on patients' thoughts and experiences on the application of technology. In the context of the rapid expansion of technology, it is important to have more detailed information on the user experience so that subsequent developers can have a clearer direction, and at the same time, have the most relevant technological tools for the core users (ie, older adult patients).

Strengths

The main advantage of the robot developed in this study is that the prototype was developed using data from professional team meetings, literature, real-world data, and experience interviews, as well as the participation of people from the relevant professional fields. This comprehensive approach ensured the prototype is robust. In addition, the development of the prototype robot was not a one-off exercise. The aim was to create a prototype that would enable actual interaction with patients with diabetes, not just from the developer's perspective. Each time further progress is made, it will be compared and revised with clinical practitioners to create a more realistic diabetes care robot that can actually interact with patients with diabetes in the real world.

Limitations

The literature reviewed in this study primarily comes from Western countries, which may not fully correspond to the disease characteristics of East Asian populations. Differences in illness, age, and cultural backgrounds may affect users' perspectives. To mitigate this limitation, future studies could be conducted in Western countries to validate the findings and evaluate their applicability in different contexts. In addition, the testing site for this study was a community pharmacy in an urban area, so it does not provide insights into how elderly individuals with different lifestyles or from different regions perceive and use the robot. Consequently, the findings may not be fully applicable to all health care settings. Future research could extend to rural community pharmacies or other community health care settings, including community hospitals, clinics, and chain pharmacies, to test the application of health care robots. Moreover, the size and functionality of the robots used in this study, along with the critical role of the internet in their operation, may limit their capabilities. For example, uninterrupted and smooth conversations rely on a stable network connection. Interruptions or network issues can hinder the robot's performance. Efforts were made to ensure a stable network environment during testing, but future studies should consider testing in various network conditions to better understand the impact of connectivity on robot functionality.

Conclusion

The prototype diabetes care robot was developed using data from the literature, professional team meetings, and practical interviews, with the aim of creating a more evidence-based health care robot that can actually give patients with diabetes a realistic experience for their disease care and enable them to give clearer and more specific feedback for the subsequent development and application of technology in disease care.

Acknowledgments

This work was supported by a grant from the Taiwan Ministry of Science and Technology (MOST 108-2634-F-006 -006-). The Taiwan Ministry of Science and Technology had no role in the design of the study, data collection, data analysis, data interpretation, or writing of the report.

Authors' Contributions

Study concept and design: CJC. Acquisition of data: CJC, LCH. Analysis and interpretation of data: CJC, LCH. Drafting of the manuscript: CJC, LCH. Critical revision of the manuscript for important intellectual content: CJC, LCH, CYC. All authors have read and approved the manuscript.

References

1. GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet 2023 Jul 15;402(10397):203-234. [doi: 10.1016/S0140-6736(23)01301-6] [Medline: 37356446]
2. Magliano DJ, Boyko EJ, IDF Diabetes Atlas 10th Edition Scientific Committee. IDF Diabetes Atlas, 10th edition: International Diabetes Federation; 2021. URL: https://diabetesatlas.org/atlas/tenth-edition/
3. Tomic D, Shaw JE, Magliano DJ. The burden and risks of emerging complications of diabetes mellitus. Nat Rev Endocrinol 2022 Sep;18(9):525-539. [doi: 10.1038/s41574-022-00690-7] [Medline: 35668219]
4. World Health Organization. Global Report on Diabetes: World Health Organization; 2016.

5. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes - global burden of disease and forecasted trends. J Epidemiol Glob Health 2020 Mar;10(1):107-111. [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
6. Department of Statistics MoHaW, Rebublic of China (Taiwan). Prevalence of high blood pressure, high blood sugar, and high cholesterol among individuals aged 20 and above [article in chinese]. 2025 URL: https://dep.mohw.gov.tw/DOS/cp-5339-59467-113.html#_2.%E7%96%BE%E7%97%85%E7%BD%B9%E6%82%A3
7. American Diabetes Association Professional Practice Committee. 13. Older adults: Standards of care in diabetes-2024. Diabetes Care 2024 Jan 1;47(Suppl 1):S244-S257. [doi: 10.2337/dc24-S013] [Medline: 38078580]
8. Goyal S, Cafazzo JA. Mobile phone health apps for diabetes management: current evidence and future developments. QJM 2013 Dec;106(12):1067-1069. [doi: 10.1093/qjmed/hct203] [Medline: 24106313]
9. Sneha S, Thalla S, Rischie I, Shahriar H. Health internet technology for chronic conditions: review of diabetes management apps. JMIR Diabetes 2021 Aug 31;6(3):e17431. [doi: 10.2196/17431] [Medline: 34463627]
10. IQVIA. Digital Health Trends 2021. 2021. URL: https://www.iqvia.com/insights/the-iqvia-institute/reports-and-publications/reports/digital-health-trends-2021
11. Huang R, Li H, Suomi R, Li C, Peltoniemi T. Intelligent physical robots in health care: systematic literature review. J Med Internet Res 2023 Jan 18;25:e39786. [doi: 10.2196/39786] [Medline: 36652280]
12. Food and Drug Administration MoHaW, Rebublic of China (Taiwan). Basic information of pharmacy [Article in Chinese]. 2024. URL: https://data.gov.tw/dataset/6134
13. Food and Drug Administration MoHaW, Rebublic of China (Taiwan). Community pharmacies: professional hubs where medicine keeps us healthy [article in chinese]. 2017 URL: https://www.mohw.gov.tw/cp-3558-37574-1.html
14. Müller P, Jahn P. Cocreative development of robotic interaction systems for health care: scoping review. JMIR Hum Factors 2024 Sep 12;11:e58046. [doi: 10.2196/58046] [Medline: 39264334]
15. Salmani Nodooshan H, Choobineh A, Razeghi M, Shahnazar Nezhad Khales T. Designing, prototype making and evaluating a mechanical aid device for patient transfer between bed and stretcher. Int J Occup Saf Ergon 2017 Dec;23(4):491-500. [doi: 10.1080/10803548.2016.1274161] [Medline: 28041534]
16. Anastassova M, Mégard C, Burkhardt JM. Prototype evaluation and user-needs analysis in the early design of emerging technologies. Presented at: Human-Computer Interaction Interaction Design and Usability, 12th International Conference, HCI International 2007; Jul 22-27, 2007; Beijing, China.
17. Chiu CJ, Hua LC, Chou CY, Chiang JH. Robot-enhanced diabetes care for middle-aged and older adults living with diabetes in the community: a small sample size mixed-method evaluation. PLoS ONE 2022;17(4):e0265384. [doi: 10.1371/journal.pone.0265384] [Medline: 35427359]
18. American Association of Diabetes Educators. An effective model of diabetes care and education: revising the AADE7 self-care behaviors ®. Diabetes Educ 2020 Apr;46(2):139-160. [doi: 10.1177/0145721719894903]
19. Klein R. Age-related eye disease, visual impairment, and driving in the elderly. Hum Factors 1991 Oct;33(5):521-525. [doi: 10.1177/001872089103300504] [Medline: 1769672]
20. Lee CF. Legibility of text-color combinations of PDA for the elderly. J Des 2005;10(2):1-12. [doi: 10.6381/JD.200506.0001]
21. Shams L, Seitz AR. Benefits of multisensory learning. Trends Cogn Sci 2008 Nov;12(11):411-417. [doi: 10.1016/j.tics.2008.07.006] [Medline: 18805039]
22. Wu YH, Fassert C, Rigaud AS. Designing robots for the elderly: appearance issue and beyond. Arch Gerontol Geriatr 2012;54(1):121-126. [doi: 10.1016/j.archger.2011.02.003] [Medline: 21349593]
23. Bedaf S, Marti P, De Witte L. What are the preferred characteristics of A service robot for the elderly? A multi-country focus group study with older adults and caregivers. Assist Technol 2019;31(3):147-157. [doi: 10.1080/10400435.2017.1402390] [Medline: 29125807]
24. Wang RH, Sudhama A, Begum M, Huq R, Mihailidis A. Robots to assist daily activities: views of older adults with Alzheimer's disease and their caregivers. Int Psychogeriatr 2017 Jan;29(1):67-79. [doi: 10.1017/S1041610216001435] [Medline: 27660047]
25. Liang A, Piroth I, Robinson H, et al. A pilot randomized trial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc 2017 Oct 1;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]
26. Broadbent E, Garrett J, Jepsen N, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45. [doi: 10.2196/jmir.8640] [Medline: 29439942]
27. Petersen S, Houston S, Qin H, Tague C, Studley J. The utilization of robotic pets in dementia care. J Alzheimers Dis 2017;55(2):569-574. [doi: 10.3233/JAD-160703] [Medline: 27716673]
28. Beuscher LM, Fan J, Sarkar N, et al. Socially assistive robots: measuring older adults' perceptions. J Gerontol Nurs 2017 Dec 1;43(12):35-43. [doi: 10.3928/00989134-20170707-04] [Medline: 28700074]
29. Sabanovic S, Reeder S, Kechavarzi B. Designing robots in the wild: in situ prototype evaluation for a break management robot. Journal of Human-Robot Interaction 2014;3(1):70. [doi: 10.5898/JHRI.3.1.Sabanovic]
30. Szajna B. Empirical evaluation of the revised technology acceptance model. Manage Sci 1996 Jan;42(1):85-92. [doi: 10.1287/mnsc.42.1.85]

31. Jayawardena C, Kuo IH, Unger U, et al. Deployment of a service robot to help older people. Presented at: 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2010); Oct 18-22, 2010; Taipei, Taiwan. [doi: 10.1109/IROS.2010.5649910]
32. Al-Taee MA, Kapoor R, Garrett C, Choudhary P. Acceptability of robot assistant in management of type 1 diabetes in children. Diabetes Technol Ther 2016 Sep;18(9):551-554. [doi: 10.1089/dia.2015.0428] [Medline: 27479198]
33. Blanson Henkemans OA, Bierman BPB, Janssen J, et al. Using a robot to personalise health education for children with diabetes type 1: a pilot study. Patient Educ Couns 2013 Aug;92(2):174-181. [doi: 10.1016/j.pec.2013.04.012] [Medline: 23684366]
34. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A systematic review of reviews evaluating technology-enabled diabetes self-management education and support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
35. Chomutare T, Fernandez-Luque L, Arsand E, Hartvigsen G. Features of mobile diabetes applications: review of the literature and analysis of current applications compared against evidence-based guidelines. J Med Internet Res 2011 Sep 22;13(3):e65. [doi: 10.2196/jmir.1874] [Medline: 21979293]
36. Holtz B, Lauckner C. Diabetes management via mobile phones: a systematic review. Telemed J E Health 2012 Apr;18(3):175-184. [doi: 10.1089/tmj.2011.0119] [Medline: 22356525]
37. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht UV. mHealth 2.0: experiences, possibilities, and perspectives. JMIR Mhealth Uhealth 2014 May 16;2(2):e24. [doi: 10.2196/mhealth.3328] [Medline: 25099752]
38. Goyal S, Morita P, Lewis GF, Yu C, Seto E, Cafazzo JA. The systematic design of a behavioural mobile health application for the self-management of type 2 diabetes. Can J Diabetes 2016 Feb;40(1):95-104. [doi: 10.1016/j.jcjd.2015.06.007] [Medline: 26455762]
39. Sinoo C, van der Pal S, Blanson Henkemans OA, et al. Friendship with a robot: Children's perception of similarity between a robot's physical and virtual embodiment that supports diabetes self-management. Patient Educ Couns 2018 Jul;101(7):1248-1255. [doi: 10.1016/j.pec.2018.02.008] [Medline: 29548599]
40. Pei D, Liang B, Du W, et al. Multimedia patient education to assist oral impression taking during dental treatment: A pilot study. Int J Med Inform 2017 Jun;102:150-155. [doi: 10.1016/j.ijmedinf.2017.03.016] [Medline: 28495343]
41. Causo A, Vo GT, Chen IM, Yeo SH. Design of robots used as education companion and tutor. In: Robot Mechatron (2015): Springer International Publishing; 2016.

Abbreviations

AADE: American Association of Diabetes Educators

Edited by A Kushniruk; submitted 22.04.23; peer-reviewed by D Dinh, PJ Long Guo, YL Leung; revised version received 14.01.25; accepted 14.01.25; published 18.03.25.

Please cite as:
Chiu CJ, Hua LC, Chiang JH, Chou CY
User-Centered Prototype Design of a Health Care Robot for Treating Type 2 Diabetes in the Community Pharmacy: Development and Usability Study
JMIR Hum Factors 2025;12:e48226
URL: https://humanfactors.jmir.org/2025/1/e48226
doi:10.2196/48226

A User-Centered Design Approach for a Screening App for People With Cognitive Impairment (digiDEM-SCREEN): Development and Usability Study

Michael Zeiler[1*], MSc; Nikolas Dietzel[2*], PhD; Fabian Haug[3,4], MSc; Julian Haug[3,4], MSc; Klaus Kammerer[3], MSc; Rüdiger Pryss[3,4], PhD; Peter Heuschmann[3,4,5], MD, PhD; Elmar Graessel[6], MD, PhD; Peter L Kolominsky-Rabas[2], MD, MBA, PhD; Hans-Ulrich Prokosch[1], PhD

[1]Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[2]Interdisciplinary Center for Health Technology Assessment and Public Health, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[3]Institute of Clinical Epidemiology and Biometry, Julius-Maximilians-Universität Würzburg, Würzburg, Germany

[4]Institute for Medical Data Science, University Hospital Würzburg, Würzburg, Germany

[5]Clinical Trial Centre Würzburg, University Hospital Würzburg, Würzburg, Germany

[6]Uniklinik Erlangen Department of Psychiatry and Psychotherapy, Center for Health Services Research in Medicine, Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen, Germany

[*]these authors contributed equally

Corresponding Author:
Michael Zeiler, MSc
Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Abstract

Background: Dementia is a widespread syndrome that currently affects more than 55 million people worldwide. Digital screening instruments are one way to increase diagnosis rates. Developing an app for older adults presents several challenges, both technical and social. In order to make the app user-friendly, feedback from potential future end users is crucial during this development process.

Objective: This study aimed to establish a user-centered design process for the development of digiDEM-SCREEN, a user-friendly app to support early identification of persons with slight symptoms of dementia.

Methods: This research used qualitative and quantitative methods and involved 3 key stakeholder groups: the digiDEM research team, the software development team, and the target user group (older adults ≥65 years with and without cognitive impairments). The development of the screening app was based on an already existing and scientifically analyzed screening test (Self-Administered Tasks Uncovering Risk of Neurodegeneration; SATURN). An initial prototype was developed based on the recommendations for mobile health apps and the teams' experiences. The prototype was tested in several iterations by various end users and continuously improved. The app's usability was evaluated using the System Usability Scale (SUS), and verbal feedback by the end users was obtained using the think-aloud method.

Results: The translation process during test development took linguistic and cultural aspects into account. The texts were also adapted to the German-speaking context. Additional instructions were developed and supplemented. The test was administered using different randomization options to minimize learning effects. digiDEM-SCREEN was developed as a tablet and smartphone app. In the first focus group discussion, the developers identified and corrected the most significant criticism in the next version. Based on the iterative improvement process, only minor issues needed to be addressed after the final focus group discussion. The SUS score increased with each version (score of 72.5 for V1 vs 82.4 for V2), while the verbal feedback from end users also improved.

Conclusions: The development of digiDEM-SCREEN serves as an excellent example of the importance of involving experts and potential end users in the design and development process of health apps. Close collaboration with end users leads to products that not only meet current standards but also address the actual needs and expectations of users. This is also a crucial step toward promoting broader adoption of such digital tools. This research highlights the significance of a user-centered design approach, allowing content, text, and design to be optimally tailored to the needs of the target audience. From these findings, it can be concluded that future projects in the field of health apps would also benefit from a similar approach.

(JMIR Hum Factors 2025;12:e65022) doi:10.2196/65022

KEYWORDS

dementia; usability; development; digiDEM; cognitive impairment; older adults; aging; mobile health; mHealth; design; feedback; screening; user centred; cognitive disorder; user-centered; mobile app

Introduction

Dementia is a widespread syndrome that currently affects over 55 million people worldwide, with annually almost 10 million new cases. The diagnosis of and treatment for people with dementia are going to be among the biggest challenges for health care systems worldwide [1]. A study by Eichler et al [2] found that 60% of people living with dementia in Germany had no formal diagnosis. Another problem lies in the long diagnostic periods. In the Bavarian Dementia Survey (BayDem) study, Wolff et al [3] found that the median time between the first perceived symptoms and diagnosis in Bavaria was 16 months. As Barth et al [4] were able to show, rural areas are also particularly affected here due to a high difficulty in accessing the facilities needed to diagnose and treat patients with dementia.

Screening instruments are one way of improving the diagnosis rate. A study with 146 participants has shown that diagnoses could be increased by almost 50% through upstream cognitive screening [2]. Internet-based screening tools offer the additional advantage that they can be used at a low threshold, regardless of time and place [5]. Digital technologies and the internet are already playing an increasingly central role in the everyday lives of older people. The proportion of people with internet access is growing across all age groups. In recent years, the number of German senior citizens (79 - 84 years) who regularly use the internet has more than doubled (18.8% in 2011 vs 39.4% in 2017). There is also an increased interest in health websites among older people [6]. In the 2021 report published by the German Federal Office for Information Security, it was stated that around 163,000 different health apps existed [7]. However, there is a lack of high-quality dementia apps. As analyzed in an earlier study, for only 6 of 20 identified dementia apps, scientific evaluation studies have been published. In none of those studies, the effectiveness of the respective screening app could be proven. Among the published app evaluations, screening apps received the worst overall quality rating. In summary, the analysis showed that the existing apps at this time did not provide reliable information and results [8].

Thus, in the digiDEM Bayern project (Digital Dementia Registry Bavaria), we have not only focused on the establishment of a digital registry for persons with mild cognitive impairment and mild to moderate dementia [9,10], but also on the development, scientific evaluation, and sustained provisioning of innovative eHealth tools and digital apps [11,12]. The goal of our current project, in this context, was to establish a user-centered design process for the development of digiDEM-SCREEN, a user-friendly app to support the early identification of persons with slight symptoms of dementia. The objective of this publication is to illustrate the iterative and agile user-centered development process consisting of 8 phases to move from a conceptual idea to an early prototype and a final prototypical implementation with continual involvement and feedback process from stakeholders and intended future users.

Methods

Overview

In order to achieve the goal of a user-friendly screening app for people with slight dementia symptoms, a user-centered iterative development approach comprising the following steps was chosen:

1. A systematic literature research of scientifically evaluated digital and nondigital dementia screening tests.
2. An early prototype (V1) development based on the guidelines for graphic design and textual formulation criteria for people with cognitive impairments. Graphical requirements include an easy-to-understand layout, standardized navigation elements, and a clear division of instructions into several steps [13]. In addition, textual guidelines such as short and concise sentences, logically structured sections with headings and an active approach to the user should be observed [14]. Furthermore, the expertise of 3 clinicians from different disciplines with long-term experience in dementia research and 2 professors of medical informatics with expertise in developing mobile health apps, supported by their teams was incorporated into the development.
3. Conduction of an initial evaluation of the early prototype (V1) based on a focus group discussion (FGD) with potential end users (older adults ≥65 years with and without subjective cognitive impairments) [15]. The group discussion was recorded and transcribed afterward. The results were then categorized and analyzed based on a previously published qualitative content analysis [16]. The following categories were extracted: general linguistic adaptations, task-related linguistic adaptations, menu navigation, general navigation, and specific design changes to individual components.

1. Incorporating the FGD feedback and results into the specification for the prototypical implementation of digiDEM-SCREEN (V2).
2. The second evaluation with a new group of potential users (older adults ≥65 years with and without subjective cognitive impairments) was based on the think-aloud method, where participants speak their thoughts and wishes aloud during the test and are observed by a researcher who also takes notes [17].
3. Incorporating the user feedback and think-aloud evaluation results into the enhanced specification for the subsequent digiDEM-SCREEN development step (V3).
4. Conduction of an additional focus group evaluation of the improved beta version of the app (V3) with people with migration background (nonnative German speakers).
5. Development and deployment of the first ready-to-use digiDEM-SCREEN version (V4).

Recruitment of the facilities for participation (steps 4, 5, and 7) was based upon the network of research partners in the project

digiDEM. The older adults from the facilities were informed about participation options in former group meetings (informed consent). After consenting, participants were invited to take part in the respective focus groups.

In steps 3 and 5, the System Usability Scale (SUS) has been calculated for the respective prototype versions by applying the German version of the standardized SUS questionnaire [18]. The scale can take values between 0 and 100; the higher the value, the higher the user-friendliness is categorized [19]. In addition, also in step 7, a self-assessment was used to determine technology use, interest, and expertise, each on a 5-point Likert scale (1 - 'Does not apply at all'; 5 - 'Applies completely'). The participants gave their subjective assessment and considered if they could use the app on their own (on a scale of 1 to 10) of the app and the specific components [20]. Furthermore, the participants were asked to name the most considerable problems associated with the app and if they wanted to change something.

Thus, our user-centered software design and development process included qualitative and quantitative evaluation methods at 3 different stages of the development process.

Ethical Considerations

This study was approved by the Ethics Committee of the Medical Faculty of the Friedrich-Alexander Universität Erlangen-Nürnberg (application number: 20-253_1-B; August 14, 2023). Written consent was obtained prior to the user testing and focus group discussion. All participants data were pseudonymized. The list of reidentifying data was stored separately from the analyzed data, and only authorized individuals have access to them. No one was paid to test the app.

Results

Steps 1 and 2: Development of the First Prototype (V1)

Prior to developing the screening test, we conducted systematic literature research. The search terms are shown in Multimedia Appendix 1. Criteria for the decision on a suitable screening tool were the scientifically examined psychometric properties (sensitivity and specificity), the availability of being a free to use app (not paid), and the technical feasibility of a tablet/smartphone. The main source of information was the systematic reviews of Chan et al [21] as well as the specific studies of the screening tools [22,23]. The decision on the Self-Administered Tasks Uncovering Risk of Neurodegeneration (SATURN) was based upon a group discussion about the aforementioned criteria as well as the (methodological) quality of the screening tools and the underlying scientific studies in general.

The SATURN [24] proved to be a test with particular promising diagnostic values (sensitivity: 0.92; specificity: 0.88 in dementia cases vs controls) [21]. The test is usable via a tablet. Administration time is about 10 minutes, which can be especially beneficial for older adults as shorter tests might induce less fatigue and therefore be more suitable for repeat administration compared with lengthier instruments [25]. Thus, the SATURN provides the foundation for the development and validation of a German adaption of the test usable as an app via smartphone and tablet.

To date, there is no German version of the SATURN test. The translation of the English version of the SATURN into German was carried out independently by 2 research assistants from the digiDEM Bayern project (MZ and ND) using the translate-retranslate method. Apart from some general adaptations, such as the correct assignment of the users' residence, linguistic aspects were also taken into account, and the texts were adapted to the German-speaking context. In some translations, the number of letters in the word increased noticeably (eg, farm - Bauernhof). A shorter related word (field - Feld) was then used in these places. Additional instructions were developed. The test adaptations aimed to ensure that both the implementation and the evaluation could be carried out entirely by the user or the system alone. At the start of the original SATURN test, the participant was asked to read aloud the task (close your eyes) and perform it [24]. Without a handler to check the action, there could be no subsequent evaluation (What phrase did you first read from this tablet?). Therefore, the researcher chose an alternative task (tap on the yellow circle) that also involved reading and performing an action.

The final screening test consists of tasks from 6 different cognitive domains: Comprehension, Visuospatial, Orientation, Memory, Calculation, and Executive Function. Points are awarded for each task, which adds up to a maximum score of 30. The tasks must be completed without the help of other people. Participants may use their visual aids to complete the tasks; all other aids (eg, paper and pencil) are not permitted. A detailed description and illustrations of the individual test tasks can be found in Multimedia Appendix 1.

Another innovation is that the authors are developing the app as a tablet and a smartphone version. Due to the smaller display sizes, new components like the word selection task shown in Figure 1 had to be created. The researcher also developed some new logic to prevent larger adjustments, for example, that the user is only allowed to undo the last connection at the last task (tap on the circle with a blue background; Figure 2).

The following table (Table 1) shows the baseline characteristics of the participants in the usability analysis. This is followed by a description of the details of the individual events.

Figure 1. Word selection task as visible on a smartphone.

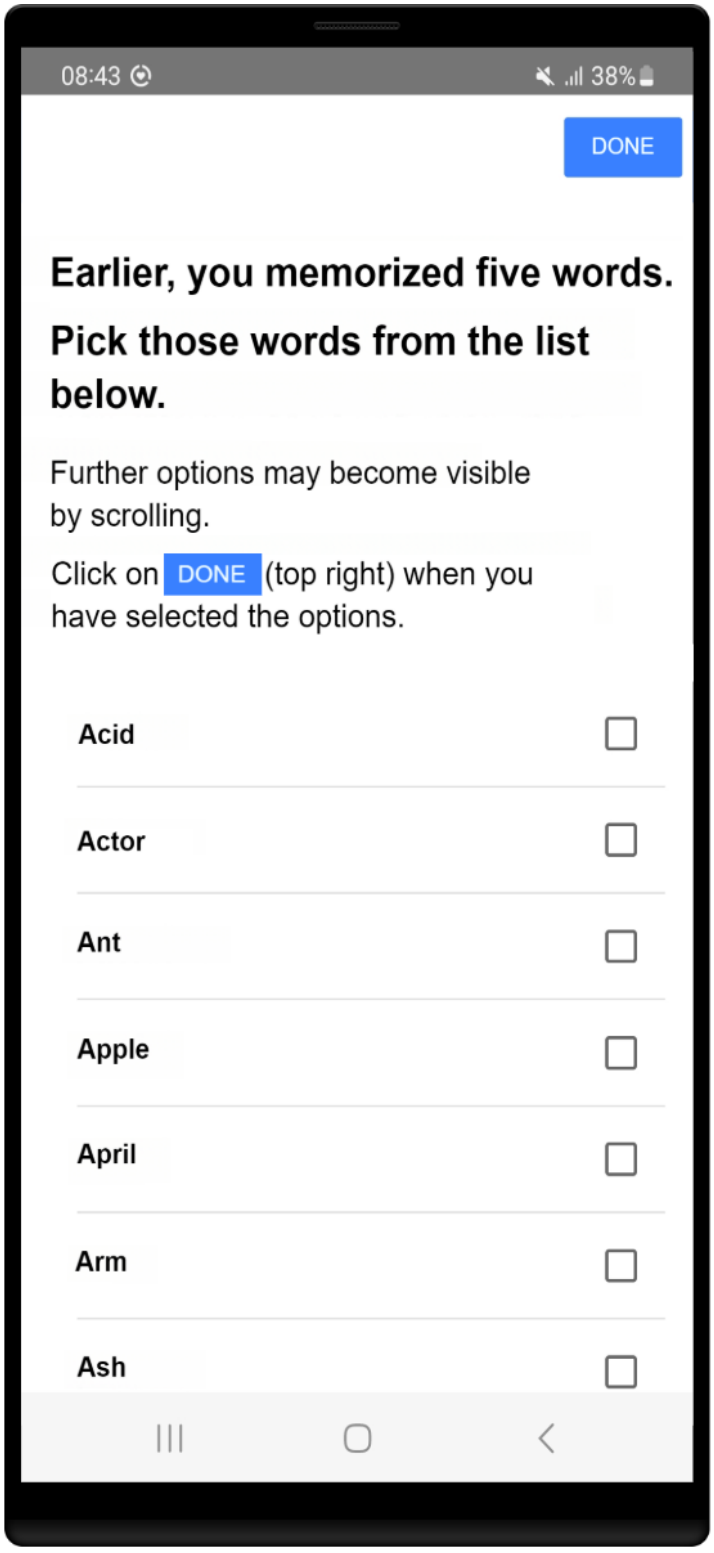


Figure 2. Trail Making Test as visible on a tablet.

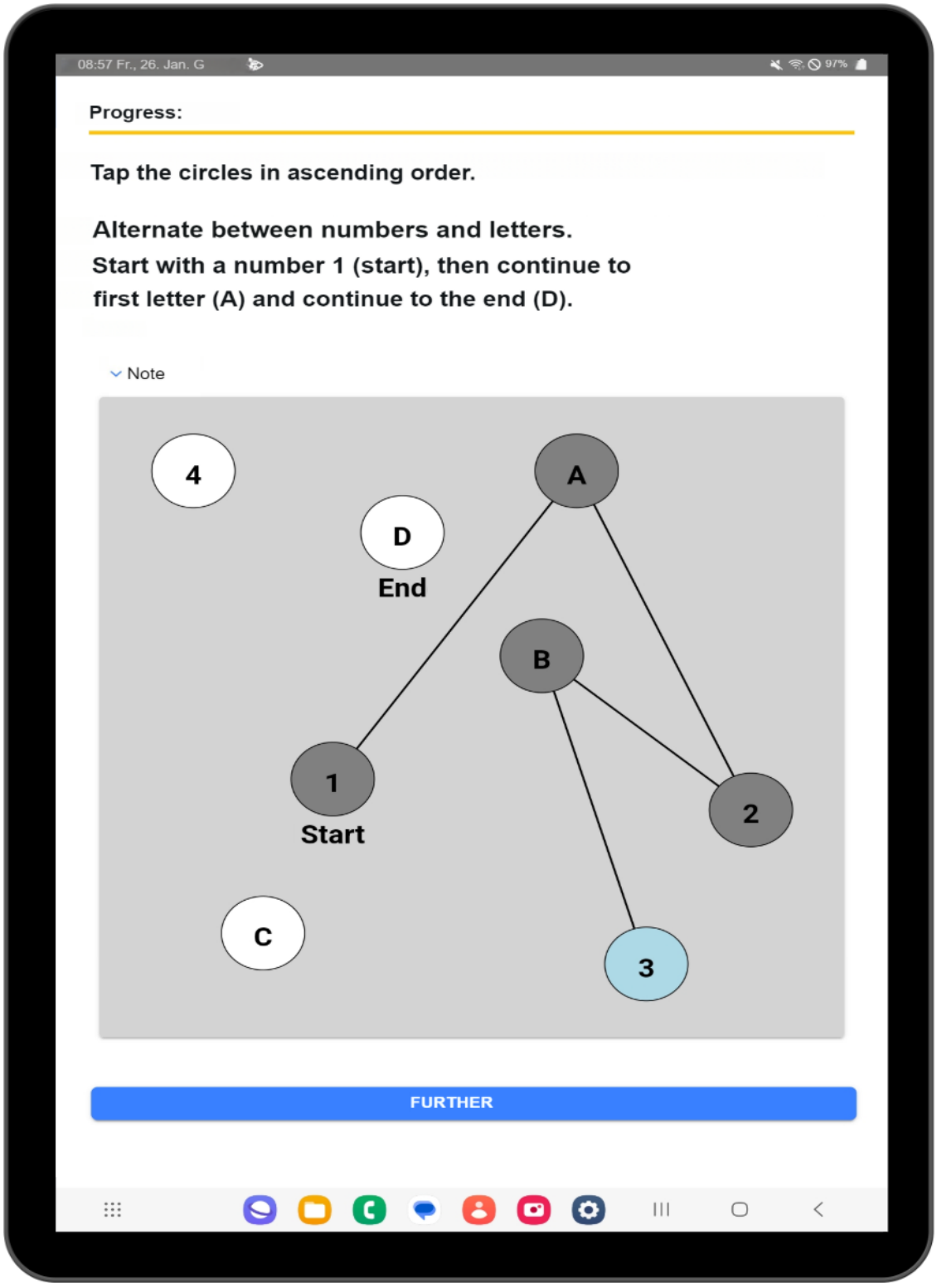


Table . Baseline characteristics.

Study sample characteristics	FGD[a] 1 (Prototype V1)	Usability test (Prototype V2)	FGD 2 (Prototype V3)
Study population	13	21	7
Age (years), mean (range)	75.8 (66-84)	70.2 (65-80)	60.8 (53-69)
Sex, n (%)			
Male	1 (7.7)	8 (38.1)	0 (0.0)
Female	12 (92.3)	13 (61.9)	7 (100.0)
Education, n (%)			
Low	4 (30.8)	0 (0.0)	1 (14.3)
Medium	5 (38.4)	10 (47.6)	1 (14.3)
High	4 (30.8)	11 (52.4)	5 (71.4)
Self-perceived cognitive impairment, n (%)	5 (38.4)	5 (23.8)	0 (0.0)
Nonnative German speaker, n (%)	0 (0.0)	2 (9.5)	7 (100.0)
SUS[b] (0 - 100), mean (SD)	72.5 (1.6)	82.4 (16.1)	—[c]
App rating (1-10), mean (SD)	7.3 (2.1)	8.7 (1.5)	8.7 (1.2)
Independent app use (1-10), mean (SD)	7.5 (2.9)	8.95 (1.3)	8.5 (1.0)

[a]FGD: focus group discussion.

[b]SUS: System Usability Scale.

[c]Not applicable.

In both the first 2 focus groups, the SUS score was slightly lower in people with subjective cognitive impairment (FGD1: healthy older adults=74.4; people with subjective cognitive impairment=69.5; FGD2: healthy older adults=83.5; people with subjective cognitive impairment=82).

The SUS score decreased with advanced age (FDG1: ≥80 years old=63.8, 79 - 70 years old=73.9, ≤70 years old=85.0; FDG2:≥80 years old=75.0, 79 - 70 years old=77.5, ≤70 years old=85.0). People with a medium-level education had the best scores on the SUS (FDG1=85.0; FDG2=83.3), followed by people with a high-level education (FDG1=73.8; FDG2=83.0) and people with a low-level education (FDG1=55.6).

Steps 3 and 4: Focus Group Discussion V1

The first FGD took place as part of a memory training group. A memory training group is a frequent meeting of older adults, in which those adults perform different memory training exercises under the supervision of a group leader. Frequent excursions are also part of this service. The service is offered by a nonprofit organization (German: Wohlfahrtsverband) and is led by a research associate in the project digiDEM. The group consisted of a total of 13 participants. Their baseline characteristics are shown in Table 1. On average, they used modern technologies frequently (3.39) and showed an average interest in technological innovations (3.00). Their self-assessment of their competence in using modern technology was moderate (2.39), but the fear of failure played only a moderately important role (2.69).

Due to the large number of participants, 3 small groups were formed for the test. Participants were able to extensively test the app prototype and contact a research assistant with any questions. The individual components of the prototype achieved a subjective app rating of 7.3 (out of 10) points and an SUS score of 72.5. Participants also generally felt able (7.5 out of 10) to use the app independently without outside help. Subsequent group discussion of the results took place again in a large group.

The 2 most significant areas of improvement were observed in all the 3 small groups. Many participants recognized the letter I as T due to the inverted commas (‘I’) and had problems answering this task correctly. Participants also did not always recognize the selected word, as only the radio button on the right-hand side of the prototype showed their selection. The app prototype (V1) is shown in Figure 3. Participants therefore specifically requested that the entire line be colored when a selection was made. The 2 adjustments are shown in Figure 4.

Figure 3. Illustration of the prototype (**V1**).

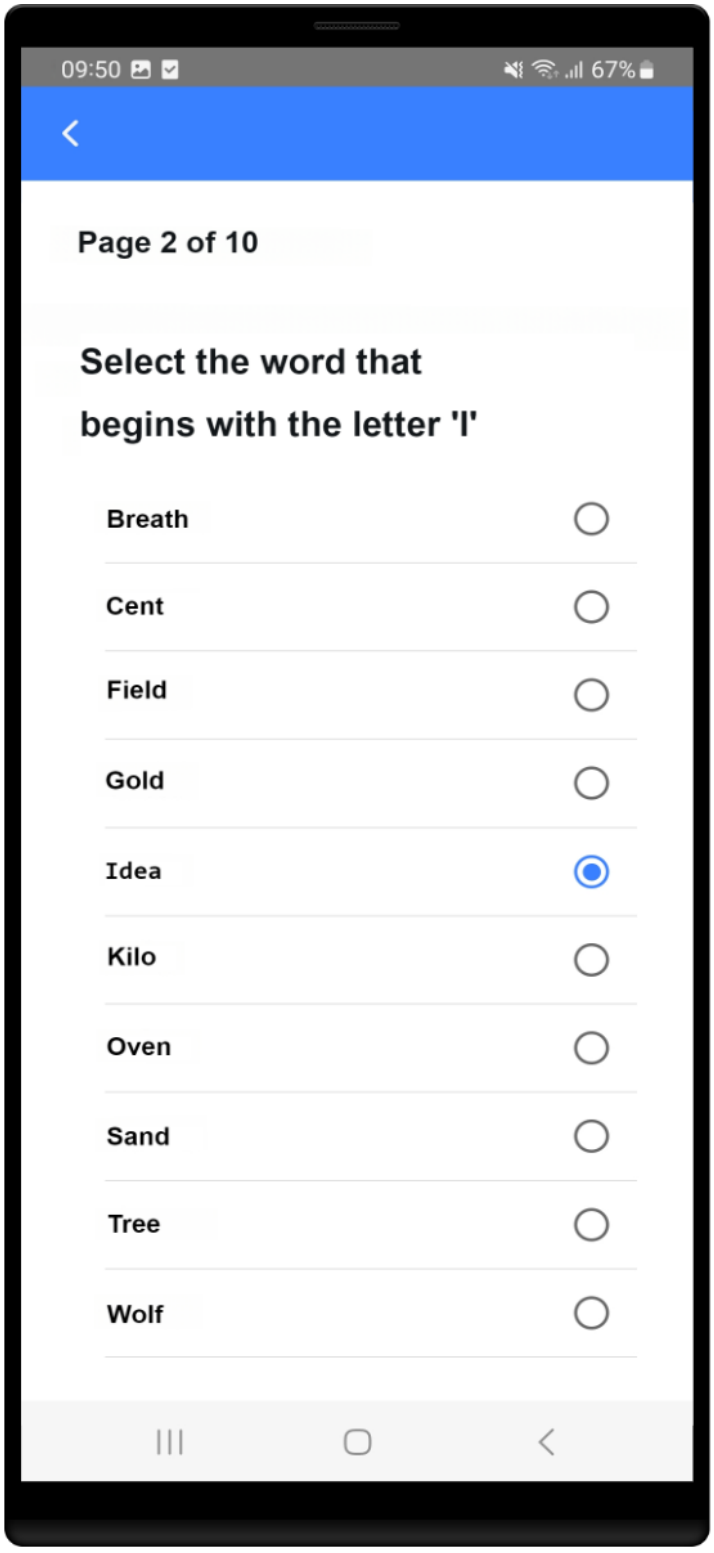
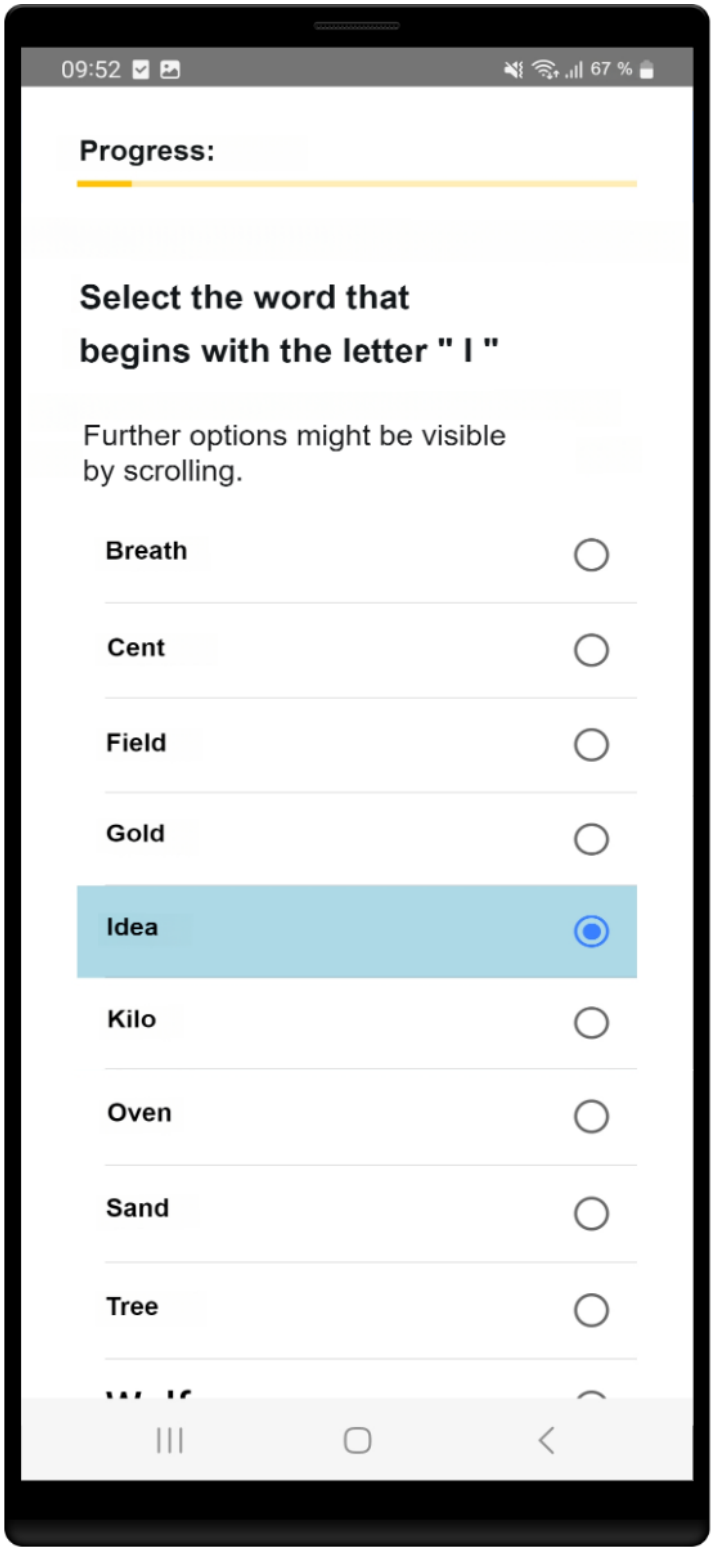


Figure 4. Final visualization (**V4**).



In addition, minor inconsistencies were noticed in this test, such as the fact that sometimes “Next” and sometimes “Done” were used to move on to the next task. The participants also wanted the selected images to be marked more clearly and the contrast and color intensity to be adjusted so that the colors could be recognized more clearly. One participant commented that she liked the “simple design” and that it did not distract from the actual content. Another participant mentioned that the instructions were too complex (“They were good instructions that you could actually understand. But I really had to read very carefully”). Therefore, the descriptive text has been simplified. Feedback on the user-friendliness of the FGD was predominantly positive. One participant particularly liked the fact that she could use the app without having much prior knowledge. The general consensus was that the app was easier to use on a tablet than on a smartphone due to the larger screen size. However, the smartphone version was also rated as usable by participants. These points mentioned were discussed with

the developers and incorporated into the second prototype accordingly.

Steps 5 and 6: Think-Aloud Usability Evaluation (V2)

A total of 21 older adults, who were randomly selected from participants in a dementia prevention event for older adults (≥65 years), participated in this usability test. Their baseline characteristics are shown in Table 1. They most frequently use modern technology (4) and are interested in technical innovations (3.62). They rated their competence in modern technology as average (3.33), while the fear of failure did not play a significant role (2.48).

The quantitative key figures collected increased compared with the first version. This prototype achieved a subjective app rating of 8.7 (out of 10) points and an SUS score of 82.4. The participants' assessment of using the app independently, without external help, also increased significantly (8.95 out of 10).

Based on the researchers' observations and the participants' statements minor adjustments and precisions, such as allowing €67 and €67.00 as the correct answer in the calculating task, were made. Some users also commented negatively about the last task's descriptive text. Due to the length and complexity of the content, the question was often not solved or only solved with a hint from the research assistant. Based on this feedback, the language of the text was revised again.

Step 7: Focus Group Discussion With Nonnative Speaker (V3)

The last user test took place under the aspect of accessible language and comprehensibility. To this end, an FGD was conducted with people with a migration background. Seven older adults took part in this FGD. Their baseline characteristics are shown in Table 1. They came from 4 different countries (Iraq, Kuwait, Sri Lanka, and Syria) and were all nonnative German speakers. They most frequently use modern technology (3.86) and are interested in technical innovations (3.57). They rated their competence in modern technology as average (3.14), while fear of failure played a minor role (2.58).

The quantitative indicators collected were similar to those of the German-speaking users. This prototype achieved a subjective app rating of 8.7 (out of 10) points. These participants also rated the success of using the app independently, without external help, at 8.5 (out of 10). Unfortunately, no SUS score could be obtained from this group due to the language barrier.

Two relevant changes emerged from the group discussion. First, a note on scrolling (Figure 4) was added in the appropriate places, and second, the language was adapted. A total of 5 of the 7 participants answered one of the initial questions (Select the fruit from the list.) incorrectly. The participants confused 2 words "Kirsche (cherry)" and "Kirche (church)," which are very similar in German. As this error does not indicate a possible cognitive decline, the word "Kirche (church)" was changed to "Kapelle (chapel)."

Step 8: Development and Deployment of the First Ready-to-Use digiDEM-SCREEN Version (V4)

In this step, the digiDEM-SCREEN test was finalized as a screening app for recording the current cognitive status of users. A validation study is currently underway. The test will be administered to patients in outpatient memory clinics and its sensitivity and specificity will be evaluated in the context of existing diagnoses and other nondigital cognitive tests. As part of the validation, cut-off values for categorizing current cognitive ability will also be determined as part of the validation. Depending on the test result, the user is given a short recommendation and options for action. If the result is above the threshold value calculated in the validation study, the screening does not indicate memory impairment. It is recommended that the test be repeated at regular intervals to monitor changes in memory performance. If the final result is below the threshold, further neuropsychological assessment in a memory outpatient clinic is recommended.

A study is currently underway to determine the sensitivity and specificity of the developed screening test (V4) and its correlation with the Montreal Cognitive Assessment [26]. For increased transparency, the research group has registered the project in the German Clinical Trials Register (DRKS) (registration number: DRKS00033764). After validation, the test will be available free of charge to anyone interested. Different randomization options have been used to minimize the learning effect. There are 5 versions of the test, which differ in the order of the numbers to be memorized. In addition, the position of each answer option is randomized for each test session. There are also plans to offer the screening test in different languages in the future. The possibility of multilingualism has already been taken into account in the programming of the app. This will be easy to implement once further translations have been validated.

Discussion

Principal Findings

There is a lack of evidence in the field of freely accessible apps for people with dementia, especially screening apps. A study published in 2023 showed that there are not any scientific studies to prove the effectiveness of any of the German-speaking screening apps [8].

A user-friendly app should have 3 main characteristics: typography appropriate for the target group (eg, recognizable icons), intuitive operation, such as fewer clicks to the desired action, and simplicity (eg, simple navigation) [27]. Three key stakeholder groups were involved in developing the digiDEM-SCREEN app: the digiDEM research team, the software development team, and the target user group (older adults ≥65 years with and without subjective cognitive impairments). Developing an app for older adults presents several challenges, both technical and social. Older adults may have less technology experience and difficulty understanding complex user interfaces [28]. Many older adults also have age-related limitations, such as visual or hearing impairments [29]. Cognitive abilities can decline with age, making complex

apps more difficult to use. The app should be tailored to the cognitive needs of older users, for example, by providing clear instructions and simple interactions. Considering these challenges when developing an app for older adults can help to create a user-friendly and accessible application that improves the lives of older people and promotes their independence [27]. In order to make the app as user-friendly as possible, feedback from potential future end users is essential. A critical examination of the study population shows that the participant structure is dominated by women in terms of gender. This could lead to a distortion of the results, as the findings may not be transferable to the entire target group. However, an empirical comparison of the usability of a mail app between male and female users showed that there are no statistically significant differences in the performance criteria of efficiency, effectiveness, and satisfaction between the 2 groups [30]. Another study examined whether there were systematic differences between women and men in the evaluation of the user experience of 3 websites and showed that there were no significant differences between the genders. Personal attitudes and preferences have a greater influence on the results [31].

The activities summarized under the term “patient and public involvement” enable patients to be actively involved in the planning and development of new products. International associations such as Alzheimer Europe as well as scientists are very interested in encouraging the active involvement of people with dementia in research for brainstorming and counseling [32].

Digital health apps often struggle with low adherence. One possible reason for this is users’ personal frustration with the content of the intervention, the way it is presented, and the nonintuitive handling [33]. The selected user-centered design could sustainably increase adherence. Users have unique knowledge, perspectives, and experiences that can influence a product’s quality, appropriateness, and user-friendliness. User testing is an essential part of the iterative development process and contributes to increasing the quality and success of the app [34]. Prototypes make it possible to recognize potential problems or weaknesses in user interaction or design at an early stage. By discovering these problems at an early stage, expensive changes or new developments in later phases of development can be avoided [35]. Thus, using the prototype design in the first workshop provided the team with a cost-effective way to get feedback and evaluate the idea. In the user tests, the tablet prototype of the app performed better than the smartphone version. The participants mainly criticized the smaller font and display size, which made it somewhat difficult to enter answers in some places. Despite these criticisms, the smartphone version was still rated as user-friendly. Smartphones are the most common mobile devices. A study by Weber et al [36] from 2020 found that an average of 41.4% of 71.6-year-old participants used their own smartphone. Due to the high availability of smartphones among seniors, the research team decided to stick with both versions. The mobile app also works offline, so no internet connection is necessary. Those decisions made it possible to reach a larger number of potential users. During the test phases, the participants did not use their devices, which they were familiar with in everyday life, but devices provided by the research team. For example, the display size, operating system (or at least the version), and individual settings may differ from their device. It is expected that user-friendliness will be even higher when users utilize their own smartphone or tablet.

The workshop participants were positive about the experience and gave constructive comments on the app. In addition, the SUS score also increased with each iteration of the app version. Bangor et al [37] described that products with a SUS score of 90 points and above were rated as exceptional, products with a SUS score of 80 points were rated as good, and products with a SUS score of 70 points were rated as acceptable. Anything below 70 points had usability issues that were a cause for concern. That means all SUS assessments are above the average and at least rated as acceptable. The prototype V1 gained a SUS score of 72.5 from the participants of the FGD. The rating of the second app version was notably better (SUS score: 82.4) and, therefore, rated as good. Due to some comprehension difficulties, unfortunately, no SUS score could be raised in the second FGD. Whenever questionnaires are used directly with people with dementia, the questions should be short and understandable (no technical terms), and double negatives should be avoided [27].

The results of the user evaluations showed that a user-friendly screening test for people with subjective cognitive impairments could be developed for the German-speaking population.

The main focus of the focus groups was on minimizing potential sources of error. Nevertheless, there is still a residual risk that the test cannot be carried out properly. If the first 3 simple test tasks are not answered correctly, an end screen appears with the message that the test cannot be carried out due to technical or language barriers. In this case, the user is advised to visit a specialized clinician. In the general instructions before the start of the test, participants are also informed that for example visual aids should be used.

Strengths

Different potential end users were included in the development process of the digital screening test as an app in order to improve usability and avoid technical or linguistic barriers. Moreover, additional languages and other extensions like a dementia prevention module can be added to the app.

Limitations

There was no random sampling of participants. Furthermore, although the SUS scale is the most frequently used scale to assess the user-friendliness of IT applications, it also has its weaknesses. The results of a systematic review show that some studies found that the double-negative questions from the SUS are challenging to understand for people with dementia [27]. With this knowledge in mind, we used simple language and no double negatives for the remaining questions we phrased. In the first FGD, the researchers had to explain individual questions to the group, and no SUS score could be collected in the last FGD. Most of the participants needed help understanding the questions. They were, therefore, unable to give reliable answers.

Conclusions

The development of digiDEM-SCREEN serves as an excellent example of the importance of involving experts and potential end users in the design and development process of health apps. From the initial stages of the project, experts were engaged in the content and design realization, providing a solid foundation for further development. The intensive testing phase, in which various end users tried out the app prototype in several iterations, clearly demonstrated the value of early and continuous feedback for improving the final product. This research highlights the significance of a user-centered design approach, allowing content, text, and design to be optimally tailored to the needs of the target audience. From these findings, it can be concluded that future projects in the field of health apps would also benefit from a similar approach. Research teams and app developers should integrate user-centered design practices into their development processes to ensure that the applications they create are not only functional but also user-friendly and appealing to the target audience. Such an approach could significantly enhance the acceptance and effectiveness of health apps, thereby making a valuable contribution to digital health care. A close collaboration with end users leads to products that not only meet current standards but also address the actual needs and expectations of users. This is a crucial step toward improving health technology and promoting broader adoption of such digital tools.

Acknowledgments

The present work was performed by MZ in partial fulfillment of the requirements for the degree “Dr. rer. biol. hum.” at the Faculty of Medicine of the Friedrich-Alexander-Universität Erlangen-Nürnberg. The project is funded by the Bavarian State Ministry of Health, Care and Prevention (funding code: G42d-G8300-2017/1606-83).

Authors' Contributions

MZ, ND, RP, PH, PKR, and HUP contributed to the conceptualization of the study. MZ and ND conducted the investigation, while MZ, ND, and HUP developed the methodology. Project administration was managed by RP, PH, EG, PKR, and HUP. The app programming was carried out by FH, JH, KK, and RP. Supervision was provided by MZ, ND, PKR, and HUP. MZ was responsible for writing the original draft, and MZ, NDI, EG, KK, and HUP contributed to the review and editing of the manuscript.

Conflicts of Interest

RP is a partner in Lenox UG, a company focused on applying scientific findings in digital health applications. Lenox UG holds shares in HealthStudyClub GmbH. This company is responsible for developing the app presented in the paper. RP received consultancy fees, reimbursements for conference attendance, and travel expenses in connection with the topics of mobile health and e-mental health. KK is also a shareholder in HealthStudyClub GmbH and has joined the company as chief technical developer. FH and JH work as freelancers for HealthStudyClub GmbH.

Multimedia Appendix 1
Search terms and detailed test description.
[DOCX File, 2717 KB - humanfactors_v12i1e65022_app1.docx]

References

1. Global status report on the public health response to dementia. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240033245 [accessed 2024-06-15]
2. Eichler T, Thyrian JR, Hertel J, et al. Rates of formal diagnosis of dementia in primary care: the effect of screening. Alzheimers Dement (Amst) 2015 Mar;1(1):87-93. [doi: 10.1016/j.dadm.2014.11.007] [Medline: 27239495]
3. Wolff F, Dietzel N, Karrer L, et al. Timely diagnosis of dementia: results of the Bavarian Dementia Survey (BayDem). Gesundheitswesen 2020 Jan;82(1):23-29. [doi: 10.1055/a-1031-9559] [Medline: 31863447]
4. Barth J, Nickel F, Kolominsky-Rabas PL. Diagnosis of cognitive decline and dementia in rural areas - a scoping review. Int J Geriatr Psychiatry 2018 Mar;33(3):459-474. [doi: 10.1002/gps.4841] [Medline: 29314221]
5. Labrique A, Vasudevan L, Mehl G, Rosskam E, Hyder AA. Digital health and health systems of the future. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):S1-S4. [doi: 10.9745/GHSP-D-18-00342] [Medline: 30305334]
6. Eighth report on older people and digitalisation - statement by the federal government (parliamentary paper). Bundesministerium für Familie Senioren, Frauen und Jugend (BMFSFJ). URL: https://www.bmfsfj.de/resource/blob/159916/9f488c2a406ccc42cb1a694944230c96/achter-altersbericht-bundestagsdrucksache-data.pdf [accessed 2024-06-15]
7. IT security in the digital consumer market: focus on health apps. Bundesamt für Sicherheit in der Informationstechnik. 2021. URL: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/DVS-Berichte/gesundheitsapps.pdf?__blob=publicationFile&v=2 [accessed 2024-06-15]
8. Zeiler M, Chmelirsch C, Dietzel N, Kolominsky-Rabas PL. Scientific evidence and user quality in mobile health applications for people with cognitive impairments and their caregivers. Z Evid Fortbild Qual Gesundhwes 2023 Apr;177:10-17. [doi: 10.1016/j.zefq.2023.01.003] [Medline: 36890031]

9. Dietzel N, Kürten L, Karrer L, et al. Digital dementia registry Bavaria-digiDEM Bayern: study protocol for a multicentre, prospective, longitudinal register study. BMJ Open 2021 Feb 8;11(2):e043473. [doi: 10.1136/bmjopen-2020-043473] [Medline: 33558357]
10. Reichold M, Dietzel N, Chmelirsch C, Kolominsky-Rabas PL, Graessel E, Prokosch HU. Designing and implementing an IT architecture for a digital multicenter dementia registry: digiDEM Bayern. Appl Clin Inform 2021 May;12(3):551-563. [doi: 10.1055/s-0041-1731286] [Medline: 34134149]
11. Reichold M, Dietzel N, Karrer L, Graessel E, Kolominsky-Rabas PL, Prokosch HU. Stakeholder perspectives on the key components of a digital service platform supporting dementia - digiDEM Bayern. Stud Health Technol Inform 2020 Jun 23;271:224-231. [doi: 10.3233/SHTI200100] [Medline: 32578567]
12. Reichold M, Selau M, Graessel E, Kolominsky-Rabas PL, Prokosch HU. eHealth interventions for dementia - using WordPress plugins as a flexible dissemination for dementia service providers. Stud Health Technol Inform 2021 May 7;279:1-9. [doi: 10.3233/SHTI210081] [Medline: 33965911]
13. Web Content Accessibility Guidelines (WCAG) 2.1. World Wide Web Consortium. 2022. URL: https://outline-rocks.github.io/wcag/translations/WCAG21-de/#unterscheidbar [accessed 2024-10-18]
14. BMAS - easy language - a guide. Bundesministerium für Arbeit und Soziales. 2014. URL: https://www.bmas.de/DE/Service/Publikationen/Broschueren/a752-leichte-sprache-ratgeber.html [accessed 2024-10-18]
15. Data collection methods for program evaluation: focus groups. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/healthyyouth/evaluation/pdf/brief13.pdf [accessed 2024-06-15]
16. Mayring P, Fenzl T. Qualitative Inhaltsanalyse: Springer Fachmedien Wiesbaden; 2019.
17. Mey G, Mruck K. Handbuch Qualitative Forschung in Der Psychologie: Wiesbaden: VS Verlag für Sozialwissenschaften; 2010, Vol. 1.
18. Gao M, Kortum P, Oswald FL. Multi-language toolkit for the system usability scale. Int J Hum-Comput Interact 2020 Dec 13;36(20):1883-1901. [doi: 10.1080/10447318.2020.1801173]
19. Brooke J. SUS -- a quick and dirty usability scale. Usability Eval Ind 1996;189.
20. Neyer FJ, Felber J, Gebhardt C. Kurzskala technikbereitschaft (TB, technology commitment). In: Zusammenstellung Sozialwissenschaftlicher Items Und Skalen (ZIS): ZIS - GESIS Leibniz Institute for the Social Sciences; 2016. [doi: org/10.6102/zis244]
21. Chan JYC, Yau STY, Kwok TCY, Tsoi KKF. Diagnostic performance of digital cognitive tests for the identification of MCI and dementia: a systematic review. Ageing Res Rev 2021 Dec;72:101506. [doi: 10.1016/j.arr.2021.101506] [Medline: 34744026]
22. Koo BM, Vizer LM. Mobile technology for cognitive assessment of older adults: a scoping review. Innov Aging 2019 Jan;3(1):igy038. [doi: 10.1093/geroni/igy038] [Medline: 30619948]
23. García-Casal JA, Franco-Martín M, Perea-Bartolomé MV, et al. Electronic devices for cognitive impairment screening: a systematic literature review. Int J Technol Assess Health Care 2017 Jan;33(6):654-673. [doi: 10.1017/S0266462317000800] [Medline: 28920567]
24. Bissig D, Kaye J, Erten-Lyons D. Validation of SATURN, a free, electronic, self-administered cognitive screening test. Alzheimers Dement (N Y) 2020;6(1):e12116. [doi: 10.1002/trc2.12116] [Medline: 33392382]
25. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb;30(1):13-28. [doi: 10.1177/1533317514522852] [Medline: 24526761]
26. Nasreddine ZS, Phillips NA, Bédirian V, et al. The Montreal cognitive assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
27. Ye B, Chu CH, Bayat S, Babineau J, How TV, Mihailidis A. Researched apps used in dementia care for people living with dementia and their informal caregivers: systematic review on app features, security, and usability. J Med Internet Res 2023 Oct 12;25:e46188. [doi: 10.2196/46188] [Medline: 37824187]
28. Deutscher Ethikrat. Big Data and Health Report on the Public Consultation of the German Ethics Council: Deutschen Ethikrat; 2018. URL: https://www.ethikrat.org/fileadmin/Publikationen/Studien/befragung-big-data-und-gesundheit.pdf [accessed 2024-06-15]
29. Stevens G, Flaxman S, Brunskill E, Global Burden of Disease Hearing Loss Expert Group, et al. Global and regional hearing impairment prevalence: an analysis of 42 studies in 29 countries. Eur J Public Health 2013 Feb;23(1):146-152. [doi: 10.1093/eurpub/ckr176] [Medline: 22197756]
30. Weichbroth P. An empirical study on the impact of gender on mobile applications usability. IEEE Access 2022;10:119419-119436. [doi: 10.1109/ACCESS.2022.3219421]
31. Aufderhaar K, Schrepp M, Thomaschewski J. Do women and men perceive user experience differently? IJIMAI 2019;5(6):63. [doi: 10.9781/ijimai.2019.03.005]
32. Gove D, Diaz-Ponce A, Georges J, et al. Alzheimer Europe's position on involving people with dementia in research through PPI (patient and public involvement). Aging Ment Health 2018 Jun;22(6):723-729. [doi: 10.1080/13607863.2017.1317334] [Medline: 28513210]

33. Kernebeck S, Busse TS, Ehlers JP, Vollmar HC. Adherence to digital interventions in healthcare: definitions, methods and open questions. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2021 Oct;64(10):1278-1284. [doi: 10.1007/s00103-021-03415-9]
34. Beckert B, Bratan T, Friedewald M, et al. The New Role of Users in a Turbulent Innovation Environment: Karlsruhe:Fraunhofer ISI; 2021.
35. Bähr B. Prototyping Requirements. In: Prototyping of User Interfaces for Mobile Applications T-Labs Series in Telecommunication Services: Springer; 2017. [doi: 10.1007/978-3-319-53210-3_3]
36. Weber W, Reinhardt A, Rossmann C. Lifestyle segmentation to explain the online health information-seeking behavior of older adults: representative telephone survey. J Med Internet Res 2020 Jun 12;22(6):e15099. [doi: 10.2196/15099] [Medline: 32530433]
37. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4:114-123.

Abbreviations

BayDem: Bavarian Dementia Survey
FGD: focus group discussion
SATURN: Self-Administered Tasks Uncovering Risk of Neurodegeneration
SUS: System Usability Scale

Edited by A Kushniruk, D Gooch; submitted 02.08.24; peer-reviewed by M Bennasar, S Kröner; revised version received 06.11.24; accepted 03.12.24; published 22.01.25.

Please cite as:
Zeiler M, Dietzel N, Haug F, Haug J, Kammerer K, Pryss R, Heuschmann P, Graessel E, Kolominsky-Rabas PL, Prokosch HU
A User-Centered Design Approach for a Screening App for People With Cognitive Impairment (digiDEM-SCREEN): Development and Usability Study
JMIR Hum Factors 2025;12:e65022
URL: https://humanfactors.jmir.org/2025/1/e65022
doi:10.2196/65022

Effectiveness of Using a Digital Wearable Plantar Pressure Device to Detect Muscle Fatigue: Within-Subject, Repeated Measures Experimental Design

Fu-Yu Chen, PhD; Tzu-Yao Lin; Yi-Cheng Huang, MSc; Evina Widianawati, MEd

Department of Biomedical Engineering, Chung Yuan Christian University, No. 200, Zhongbei Road, Zhongli District, Toayuan City, Taiwan

Corresponding Author:
Fu-Yu Chen, PhD
Department of Biomedical Engineering, Chung Yuan Christian University, No. 200, Zhongbei Road, Zhongli District, Toayuan City, Taiwan

Abstract

Background: Muscle fatigue, characterized by reduced force generation during repetitive contractions, impacts older adults doing daily activities and athletes during sports activities. While various sensors detect muscle fatigue via muscle activity, biochemical markers, and kinematic parameters, a real-time wearable solution with high usability remains limited. Plantar pressure monitoring detects muscle fatigue through foot loading changes, seamlessly integrating into footwear to improve the usability and compliance for home-based monitoring.

Objective: This study aimed to investigate the effects of muscle fatigue on plantar pressure measurements using a self-developed wearable plantar pressure system.

Methods: Twelve healthy participants completed a 5-minute calf muscle fatigue protocol. The plantar pressures and surface electromyography (sEMG) activity of the gastrocnemius muscles were recorded before and after exercise. The plantar pressures at 6 regions and the median frequency (MDF) of sEMG were analyzed to quantify fatigue.

Results: The self-developed foot pressure system showed a significant decrease in plantar pressure peak values at the heel of the left (P=.003) and right feet (P=.001) and at the lateral toe of the left (P=.001) and right feet (P=.026). A significant increase was observed at the metatarsal head of both the left foot (P=.001) and the right foot (P=.017). The MDF of sEMG signals significantly decreased in the left (P=.001) and right gastrocnemius (P<.001).

Conclusions: Plantar pressure changes and sEMG signals effectively detect gastrocnemius muscle fatigue using the proposed wearable system, supporting the development of a wearable solution for detecting muscle fatigue suitable for home-use.

(JMIR Hum Factors 2025;12:e65578) doi:10.2196/65578

KEYWORDS

muscle fatigue; plantar pressure sensors; wearable devices; home-based monitoring

Introduction

Muscle fatigue, characterized by the reduced ability to generate adequate force during repetitive contractions [1], affects performance across different groups and activities including daily activities of older adults as well as athletes during sports. Muscle fatigue disrupts the neuromuscular system and affects different aspects such as muscle strength, force production, and movement patterns, resulting in stiffness of the muscles and joints, impaired motor control, and poor balance [2], thereby increasing the risk of injuries. For instance, in badminton players, muscle fatigue impairs neuromuscular coordination, by reducing input from the foot sole receptors, and muscle force production, leading to weaker responses and decreased reaching distance performance [3]. Running-induced fatigue reduces lower limb muscle activity, decreases shock absorption capacity, and alters plantar pressure distribution [4], thus increasing the risk of running-related injuries such as stress fractures in the foot [5]. Mello et al highlighted how fatigue delays the gastrocnemius muscle's activation by 1 second relative to the center of pressure, thus impairing balance [6]. Furthermore, Morrison et al found that muscle fatigue in older adults (aged 60 - 79 years) significantly increased reaction times, postural sway, and fall risk compared to younger individuals (aged 30 - 59 years) [7].

Muscle fatigue can be detected through various methods, including blood lactate concentration [8], electromyography (EMG) [9], mechanomyography (MMG) [10], near-infrared spectroscopy (NIRS) [11], and kinematic parameters using inertial measurement units (IMUs) [12]. While the lactate concentration provides an estimation of global fatigue, it cannot monitor fatigue in real-time [8]. MMG, NIRS, and IMUs can be used in wearable forms; however, the use of NIRS and MMG has challenges owing to issues such as time lags, the use of

MMG has susceptibility to motion artifacts [13], and the use of IMUs needs further research regarding the relationship between kinematic parameters and muscle fatigue [14]. Surface electromyography (sEMG) is a widely used wearable method for real-time monitoring of local muscle fatigue by measuring myoelectric activity [15]. Fatigue reduces muscle fiber conduction velocity, shifting the sEMG power spectrum to lower frequencies (eg, decreased median frequency) and increasing signal amplitude. sEMG captures these changes and quantifies muscle fatigue using time-domain or spectral parameters [8]. However, the use of sEMG in a home setting is difficult, owing to the challenges of correctly attaching the electrodes and regularly wearing the device without assistance [16].

Adherence is a critical factor for effective daily monitoring, as meaningful and continuous data are required for accurate analysis [17]. Wearable devices integrated with mobile apps present promising healthcare solutions for home-based monitoring. The integration of hardware sensors and software mobile apps enables continuous and unobtrusive monitoring, thereby providing real-time data analysis that can support timely interventions [18]. A suitable alternative wearable approach for muscle fatigue monitoring is the use of plantar pressure sensors, which can be seamlessly integrated into a shoe's insole for daily comfortable wear. Wearable plantar pressure sensors in the foot insole are widely used for the detection of diabetic foot ulcers [19], gait analysis, and the measurement of the ground reaction force [20] and the center of pressure [6]. Muscle fatigue impacts body mechanics and alters foot loading patterns, measurable through plantar pressure at different foot regions. Considerable research has explored changes in plantar pressure following various physical activities linked to lower limb muscle fatigue, such as badminton [21], walking [22], and long-distance walking [23].

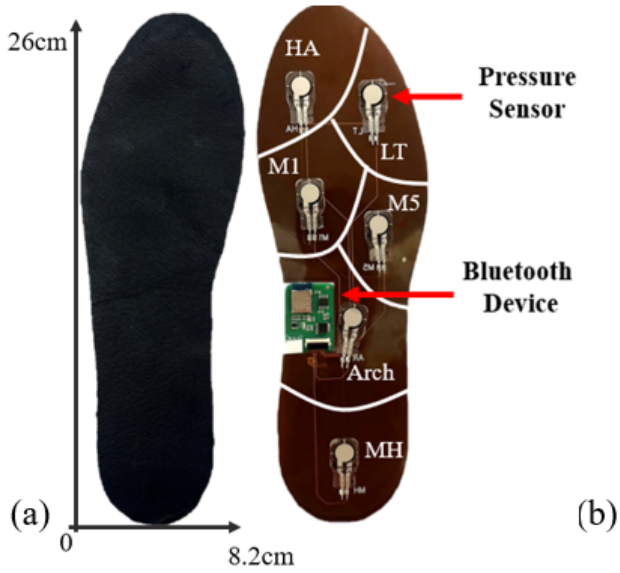
Despite the aforementioned advancements, wearable foot plantar systems for detecting muscle fatigue in everyday physical activities have yet to be designed. In this study, the effect of calf muscle fatigue in healthy participants was investigated using a self-developed wearable plantar pressure system that was suitable for daily use. The results can provide a muscle fatigue detection method for developing a wearable plantar pressure monitoring system

Methods

The Digital Foot Pressure System

A self-developed digital foot pressure system was utilized in this study. Figure 1 shows the wearable plantar devices for both the left and right feet as well as the measurements of the devices. Each device, designed to be worn comfortably with shoes, is equipped with 6 pressure sensors (A301 Flexiforce, Tekscan Inc.). The system also has a nRF52840 microprocessor (Nordic Semiconductor) used for the analog-to-digital converter measurement of foot plantar pressure, a wireless Bluetooth 5.0 module for data transmission, and a lithium-ion battery for power supply. The system operates with a sampling rate of 100 Hz, ensuring high-resolution data capture. The data from the 6 pressure sensors in both the left and right devices can be transmitted wirelessly in real time via Bluetooth to a custom-developed Android mobile app. This app facilitates real-time data display, analysis, and cloud storage, as shown in Figure 2. The pressure sensors are strategically arranged on the foot insole to cover 4 main plantar anatomical areas: the toes, metatarsals, arch, and heel. This general layout enables accurate measurement of natural gait; the layouts of the toes and metatarsals are further divided into 2 parts to capture plantar pressures along the mediolateral axis during walking [24].

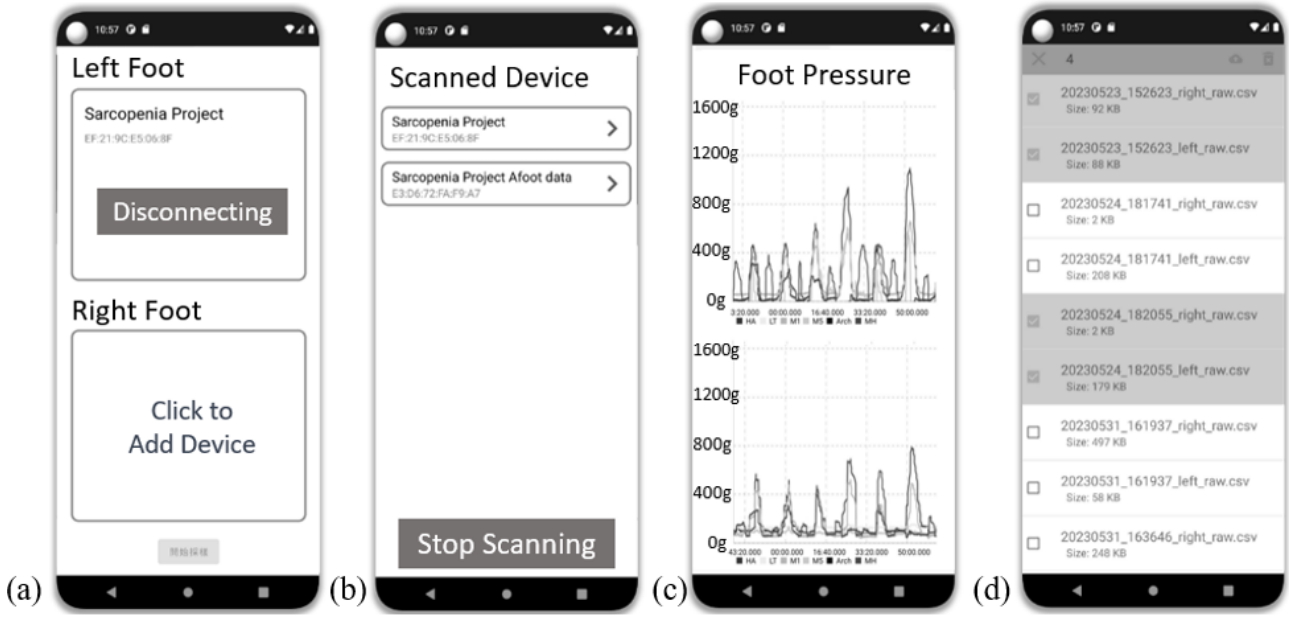
Figure 1. Wearable plantar pressure device in an insole format. (**A**) Foot insole with 6 regions of pressure measurements: hallux (HA), lesser toe (LT), first metatarsal (M1), fifth metatarsal (M5), arch, and medial heel (MH); (**B**) Coordinates of the 6 pressure sensors on the insole.



(b)

	X axis (cm)	Y axis (cm)
Hallux (HA)	5.8	21.3
Lesser Toes (LT)	2.4	21.1
First Metatarsal (M1)	6.2	16.2
Fifth Metatarsal (M5)	2.7	14.5
Arch	4.2	4.2
Medial Heel (MH)	4.7	3.9

Figure 2. Overview of the digital mobile app for the wearable plantar pressure device. (**A**) Bluetooth connection page for the devices on the right and left feet; (**B**) Bluetooth connectable device page; (**C**) real-time streaming graph showing plantar pressure data; and (**D**) Cloud storage for historical data.



Ethical Considerations

The research was approved by the Ethics Review Board of the Chung Yuan Christian University for Human Subjects Research (No. 8800-4-07-002). The participants were informed of the requirements and procedures of the entire experiment, and written consent was obtained before testing.

Subjects

A briefing session was conducted to familiarize the participants with the fatigue protocol and the walking test to be used in the research. Twelve healthy participants from the Chung Yuan Christian University, Taiwan, were recruited for the research. The mean (SD) age of the participants was 23.7 (2.6) years, and their mean (SD) weight was 62.4 (6.3) kg. The dominant limb of each participant was determined by identifying which foot would step out first once they started walking. The exclusion criteria included participants with plantar fasciitis or flatfoot and those who could not run independently for 5 minutes. Given that these foot disorders were excluded, experiments for minimizing variability could be performed. Then, suitable foot insole sizes were given to the participants.

Experimental Procedures

The experimental procedures were divided into 4 main steps. First, baseline tests were performed to measure the EMG of muscle activity and foot plantar pressure in both the legs. During the baseline test, the sEMG activities of the participants' gastrocnemius and rectus femoris muscles in the right and left legs were recorded using sEMG electrodes (T709, Comepa, France) and a data acquisition instrument (MP36, Biopac Inc., USA). The skin of each participant was cleaned with alcohol before the electrodes were attached. The sEMG muscle activities were recorded at a sampling rate of 2000 Hz. The plantar pressures at 6 locations in the left and right legs were measured simultaneously with sEMG. The plantar pressures were recorded using the wearable foot plantar device that was developed specifically for this study. Then, timed up-and-go and 10-meter walk tests were conducted; the participants were asked to rise from a standard chair, walk for 10 meters at their most comfortable speed, turn around, walk back. and sit down again [25].

After the baseline test, each participant was instructed to perform a fatigue exercise. All the participants were asked to place a weight-bearing apparatus on each leg and perform forefoot running for 5 minutes. The weight-bearing apparatus was 1/20th of a participant's weight [26]. sEMG data were collected from the gastrocnemius and rectus femoris muscles to capture the electrical activity associated with muscle contractions, and the wearable foot plantar pressure device was used to monitor changes in pressure distribution under the feet. This dual approach allowed the fatigue levels in both the muscles to be evaluated and quantified. After the exercise, the rating of perceived exertion (RPE) scale was used to assess and document each participant's perceived level of fatigue. Postfatigue measurements were recorded immediately after the fatigue protocol, while the participants performed timed up-and-go and 10-meter walk tests, and sEMG and foot plantar pressure measurements for both legs were collected.

EMG Muscle Fatigue Analysis

Three consecutive sEMG signals from the gastrocnemius and rectus femoris muscles were averaged for prefatigue and postfatigue analyses. The sEMG envelope was calculated from the raw sEMG signal through a 20 - 500–Hz finite impulse response bandpass filter. The signal was then processed via fast Fourier transformation to calculate the power spectrum median frequency (MDF) of the sEMG. A shift in the MDF to a lower frequency during exercise is considered localized muscle fatigue [27]. This method allowed researchers to determine whether muscle fatigue was induced in the gastrocnemius muscles, thereby increasing confidence in these data's correlation with foot plantar pressure data. The MDF is the frequency at which the EMG power spectrum is transmitted, and it is given by

(1)∑j=1MDF Pj=∑j=MDFM Pj=12∑j=1M pj,

where Pj is the power spectrum of the EMG at frequency bin *j*. The MDF splits the power spectrum of the EMG into two equal-amplitude regions.

The RPE Scale Analysis

The RPE scale was utilized to assess each participant's self-reported fatigue level. This subjective measurement tool is widely employed in the fields of physical activity, exercise, and sports to evaluate an individual's perceived effort and fatigue during physical tasks [28]. Participants rate their exertion based on sensations such as increased heart rate, breathing rate, muscle fatigue, and overall physical strain. The Borg RPE Scale, ranging from 6 to 20, is one of the most commonly used versions, with higher scores indicating greater levels of exertion [29].

Foot Plantar Pressure Analysis

Foot plantar pressure data, obtained using the developed digital foot plantar system and the commercial system, were simultaneously collected with the EMG data. The 6 plantar pressures in the left and right feet obtained in this research represent 6 anatomical areas of the foot: hallux (HA), lesser toe (LT), first metatarsal (M1), fifth metatarsal (M5), arch, and medial heel (MH). The peak pressures (Pa) in these 6 anatomical areas were analyzed and calculated from the raw plantar pressure data via MATLAB version R2022a (Mathworks Inc.). The measurements for the three consecutive prefatigue and postfatigue tests were averaged and compared. On the basis of the preliminary results, changes in the mean peak pressure of the 6 anatomical areas before and after the fatigue protocol were determined.

Statistical Analysis

In this study, the MDFs of the sEMG signals and the mean peak plantar pressures in each anatomical area for each participant before and after the fatigue protocol are presented as the means (standard errors). A paired *t* test with repeated measures was used to compare the MDFs of sEMG signals from the gastrocnemius and rectus femoris muscles before and after fatigue exercise. Then, a paired *t* test with repeated measures was used to detect significant differences in the 6 anatomical areas (ie, HA, LT, M1, M5, arch, and MH) before and after muscle fatigue exercise for the wearable plantar pressure system. All the statistical tests were performed with the significance level of $P<.05$.

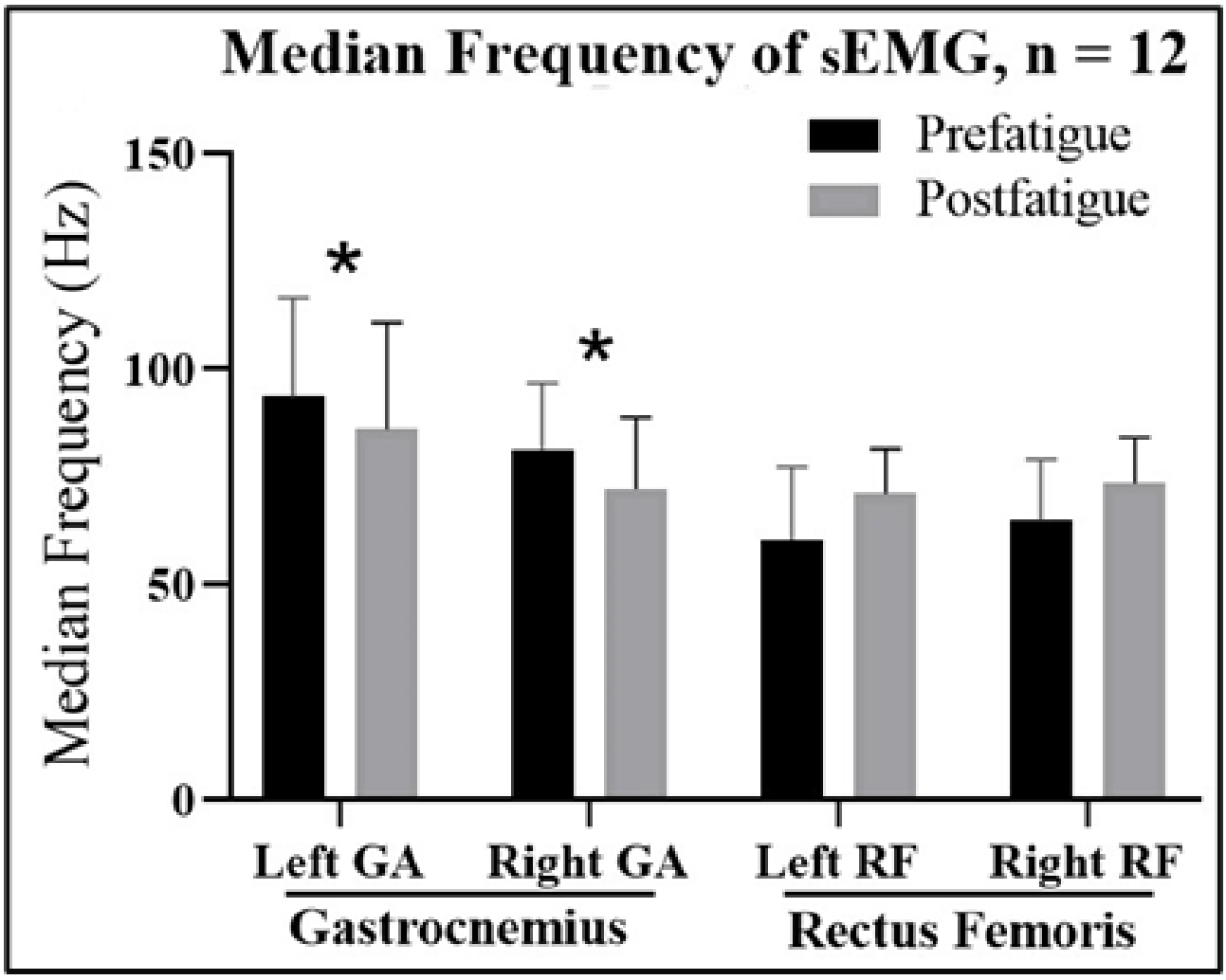
Results

The data for muscle fatigue, RPE, and mean plantar pressure were collected and analyzed using a paired *t* test.

MDFs of sEMG in Detecting Muscle Fatigue

The MDFs of the sEMG data of the gastrocnemius and rectus femoris muscles are presented in Figure 3. The MDFs (SD) of the sEMG data of the left and right gastrocnemius muscles significantly decreased from 93.6 (22.8) Hz to 86.1 (24.6) Hz in the left foot and from 81.3 (15.5) Hz to 72.1 (16.6) Hz in the right foot after exercise for muscle fatigue ($P=.001$ and $P<.001$, respectively). However, the MDF of the sEMG before and after muscle fatigue exercise did not significantly differ between the left and right rectus femoris muscles.

Figure 3. Medium frequency of the sEMG for the gastrocnemius (GA) and rectus femoris (RF) muscles in the left and right legs before the fatigue exercise (prefatigue) and after the fatigue exercise (postfatigue). sEMG data of the left and right gastrocnemius muscles significantly decreased after the exercise for muscle fatigue (*P*=.001 and *P*<.001, respectively).



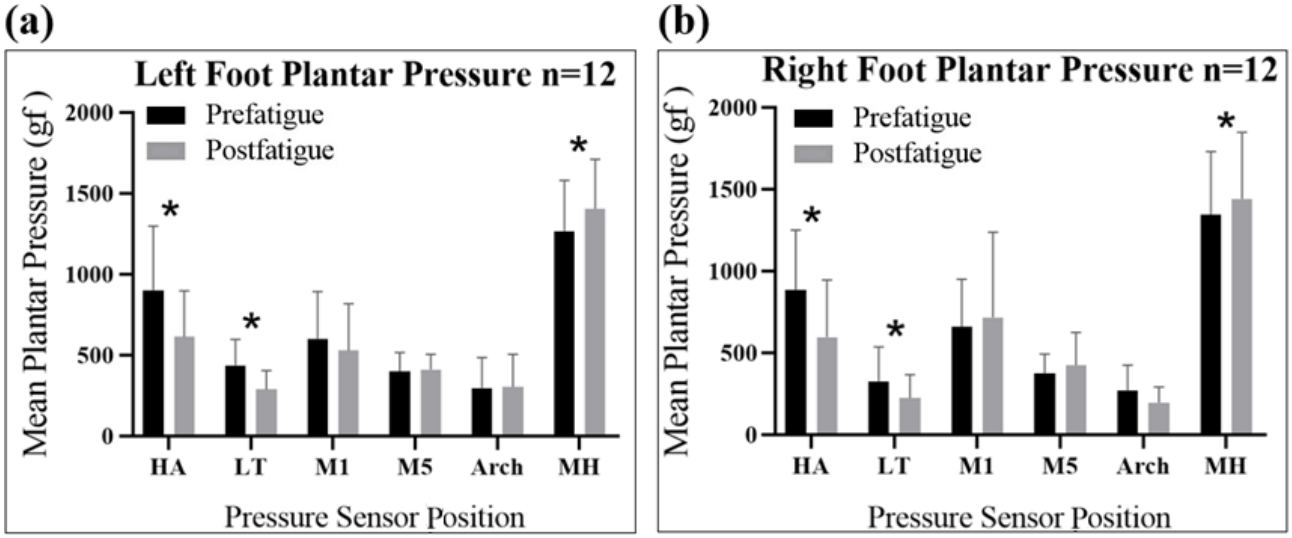
The RPE Scale in the Muscle Fatigue Protocol

The RPE scale was used in the experiment to assess the muscle fatigue levels of the 12 participants. All the participants reported that the level of fatigue in the gastrocnemius muscle was greater than that in the rectus femoris. The analysis of the average RPE scores revealed that the mean fatigue level for the gastrocnemius muscle was 14.3 (SD 1.3), whereas the mean fatigue level for the rectus femoris was 7.0 (SD 1.3). The significant difference in fatigue levels (P <.001) between the two muscles aligns with the analysis of MDF in the sEMG data. These findings confirmed that the gastrocnemius muscle experienced greater fatigue than the rectus femoris did during the experiment.

Foot Plantar Pressure in Detecting Muscle Fatigue

The data for the left and right feet (Figure 4A and B, respectively) were obtained using a developed digital plantar pressure measurement system. Figure 4 shows the changes in plantar pressure before and after fatigue. The changes were noticeable in terms of the plantar pressure at the HA, LT, and MH on both feet after gastrocnemius muscle fatigue. The pressure values at other locations also changed, but they were not statistically significant.

Figure 4. Mean peak plantar pressure before the fatigue exercise (prefatigue) and after the fatigue exercise (postfatigue) by sensor positions: hallux (HA), lesser toe (LT), first metatarsal (M1), fifth metatarsal (M5), arch, and medial heel (MH) in the (**A**) left foot: a statistically significant decrease in the mean plantar pressure peak values was observed in the HA (*P*=.003) and LT (*P*=.001), while the MH showed a statistically significant increase (*P*=.001) after the fatigue exercise; and (**B**) right foot: a statistically significant decrease in the mean plantar pressure peak values was observed in the HA (*P*=.001) and LT (*P*=.026). A significant increase was noted in the MH position after fatigue (*P*=.017).



The mean plantar pressure values were measured using the self-developed foot pressure system. A significant decrease was observed in the mean plantar pressure peak values in the HA of the left foot (*P*=.003) and right foot (*P*=.001) and in the LT of the left foot (*P*=.001) and right foot (*P*=.026). A significant increase was observed in the mean plantar pressure peak value in the MH position after fatigue in both the left and right feet (*P*=.001 and *P*=.017, respectively). The mean plantar pressure peak value in the right HA decreased by 32.6%, from 883.2 (366.8) gf to 595.1 (349.2) gf, and that in the left HA decreased by 31.9%, from 900.7 (396.2) gf to 612.8 (284.0) gf. The mean plantar pressure peak value in the right LT decreased by 31.1%, from 325.8 (210.6) gf to 223.7 (141.4) gf, and that in the left LT decreased by 33.9%, from 435.6 (162.0) gf to 287.8 (115.8) gf. The mean plantar pressure peak value in the right MH increased by 7.2%, from 1343.4 (385.7) gf to 1440.3 (406.8) gf, and that in the left MH increased by 11.0%, from 1265.5 (313.0) gf to 1404.9 (304.4) gf. The changes in the mean peak pressure in the M1, M5, and arch positions were not significant. However, the M1 and arch positions showed a change in the mean peak pressure in both the left and right feet. In contrast, we observed an increase in the mean peak pressure in the M5 position in the left and right feet of 12.5 (24.5) gf and 52.0 (79) gf, respectively. The details of the changes in the mean peak plantar pressures before and after fatigue exercise by sensor position via the wearable plantar pressure system are shown in Multimedia Appendix 1.

Discussion

Principal Findings and Comparison With Previous Works

The results of this study demonstrated that muscle fatigue in the gastrocnemius significantly affects the plantar pressure in the HA, LT, and MH regions in both the legs. In particular, the mean peak plantar pressure in the HA and LT regions significantly decreased and that in the MH regions significantly increased after 5 minutes of muscle fatigue exercise in the gastrocnemius compared with the value before the fatigue exercise. This finding was determined using the self-developed wearable foot plantar pressure system proposed in this research. Muscle fatigue exercise in this study induced muscle fatigue solely in the calf area, which corresponded with a significant decrease in the MDF of the sEMG of the gastrocnemius muscle but not in the MDF of the sEMG of the rectus femoris. As presented, sEMG is highly muscle-specific and effective for detecting localized muscle fatigue; it is not useful to detect muscle fatigue in deeper muscles like the tibialis posterior in the calf area [5]. Pressure sensors offer an alternative by measuring the fatigue in muscles through shifts in loading patterns, providing a broader and more accessible assessment.

The findings of this study align with existing literature, showing that muscle fatigue in the lower limb induces a significant decrease in the mean pressure in the HA area and a shift in plantar pressure from the forefoot to the hindfoot [21,23,30,31]. This could be due to individuals starting to adapt their plantar pressure pattern from the forefoot to the hindfoot to avoid overuse of the forefoot [23]. The literature also revealed a significant decrease in the first metatarsal region. Similar findings were observed in this study, but the results in the first and fifth metatarsals and arch regions were not statistically significant. This discrepancy could stem from differences in fatigue protocols, as more intensive activities, such as 30-minute runs, lead to greater pronation and increased medial midfoot loading, whereas walking shows no such midfoot differences [32-35]. Additionally, subject variability, including anatomical factors such as arch height and leg-length differences, may contribute to variance in medial midfoot pressure [30,36]. For instance, flat feet increase loading on the medial longitudinal arch, while high arches shift the load to the lateral edge [37].

In this study, a shift in loading from the forefoot region including the HA and LT to the hindfoot including the MH after 5 minutes of fatigue exercise results in gastrocnemius muscle

fatigue. This finding could serve as an indicator for real-time monitoring of lower limb muscle fatigue, aiding in the prevention of injuries or falls in older adults and athletes. When a healthy person participates in sports such as badminton or running, muscle fatigue of the lower limb may be expected. These fatigued muscles can result in compromised reaction times, joint stability, and dynamic balance, thereby reducing impact absorption, which heightens the risk of injuries [21,38]. Athletes and coaches should monitor for fatigue alerts and ensure timely rest to recover muscle strength, preventing performance issues and injuries [39]. For older populations, particularly those at high risk of falls, such as patients with sarcopenia [40] or stroke [41], this system enables continuous home-based monitoring of plantar pressure patterns. A shift in plantar loading from the forefoot to the hindfoot could trigger alerts for rest to individuals or caretakers, minimizing fall risks associated with muscle fatigue.

Usability is a key factor in the adoption of digital health technologies such as the proposed wearable plantar pressure system [42]. Compared to sEMG, MMG, NIRS, or IMU wearable devices, its shoe-based form offers a user-centered design for daily wear, eliminating the need for users to remember and attach it correctly, thereby enhancing acceptance and compliance in home settings. A previous study highlighted the importance of face-to-face interactions with physicians in increasing their trust in digital health technologies [43]. Clinical or expert recommendations play a critical role in encouraging patients to adopt the device. The system's ability to support effective remote patient management allows both users and experts to review and monitor past performance or clinical conditions, thus fostering engagement [44]. These features, combined with its user-friendly design and portability, enhance the system's feasibility and scalability for home-based applications, promoting widespread adoption.

Limitations

This study has some limitations. First, direct evidence for the development of muscle fatigue in the gastrocnemius could not be provided by this study, given that muscle force was not directly measured. Instead, the MDF of the sEMG signal, which shifted to a low frequency, was used as an indicator of muscle fatigue; this approach is a common research method for measuring muscle fatigue [45-47]. The decrease in MDF was likely caused by changes in the properties of the muscle fibers, such as decreased conduction velocity and increased muscle fiber recruitment [48]. The MDF of the sEMG measurement might also be affected by factors other than fatigue, such as the muscle fiber type [49] and motor unit firing rate [50]. However, this limitation was compensated by the RPE scale questionnaire. Second, the effects of fatigue were measured only in the rectus femoris and calf gastrocnemius muscles; the effect of upper leg fatigue on plantar pressure was ruled out. sEMG measurements of other calf muscles, such as the tibialis anterior, can provide insights into a particular calf muscle's interaction with plantar pressure. The short test duration and limited wearing time (10 - 20 min) in this study may not accurately represent real-life scenarios, necessitating extended testing to better simulate prolonged activities. In addition, wearable devices must adapt to various environments and wearing conditions, as factors such as daily wear and tear, temperature, humidity, flooring, and uneven terrain can impact sensor performance.

Conclusions

The results of this study demonstrated a significant decrease in the mean plantar peak pressure in the HA and LT and an increase in the mean plantar peak pressure in the MH as an indicator of the onset of muscle fatigue in the gastrocnemius. This work is an exciting proof-of-concept outcome showing that muscle fatigue in the gastrocnemius can be detected via a wearable plantar pressure system. These findings can be used to further develop a wearable lower limb muscle fatigue monitoring system for minimizing injury risk in sports or during the daily activities of older adults.

Acknowledgments

The author would like to thank the participants for their time and willingness to take part in this study. This work was supported by the Taiwan National Science and Technology Council (Grant No. 110-2222-E-033-006-MY3).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Mean peak plantar pressures before and after fatigue exercise by sensor position via the wearable plantar pressure system system.
[DOCX File, 8 KB - humanfactors_v12i1e65578_app1.docx]

References

1. Finsterer J. Biomarkers of peripheral muscle fatigue during exercise. BMC Musculoskelet Disord 2012 Nov 8;13:218. [doi: 10.1186/1471-2474-13-218] [Medline: 23136874]
2. Pethick J, Tallent J. The neuromuscular fatigue-induced loss of muscle force control. Sports (Basel) 2022 Nov 21;10(11):184. [doi: 10.3390/sports10110184] [Medline: 36422953]

3. Sarshin A, Mohammadi S, Shahraba HBP, Sedigh M. The effects of functional fatique on dynamic postural control of badminton players. Biol of Exerc 2011;7(2):25-34. [doi: 10.4127/jbe.2011.0047]
4. Apte S, Prigent G, Stöggl T, et al. Biomechanical response of the lower extremity to running-induced acute fatigue: a systematic review. Front Physiol 2021;12:646042. [doi: 10.3389/fphys.2021.646042] [Medline: 34512370]
5. Weist R, Eils E, Rosenbaum D. The influence of muscle fatigue on electromyogram and plantar pressure patterns as an explanation for the incidence of metatarsal stress fractures. Am J Sports Med 2004 Dec;32(8):1893-1898. [doi: 10.1177/0363546504265191] [Medline: 15572318]
6. Mello RGT, Oliveira LF, Nadal J. Anticipation mechanism in body sway control and effect of muscle fatigue. J Electromyogr Kinesiol 2007 Dec;17(6):739-746. [doi: 10.1016/j.jelekin.2006.08.011] [Medline: 17142060]
7. Morrison S, Colberg SR, Parson HK, et al. Walking-induced fatigue leads to increased falls risk in older adults. J Am Med Dir Assoc 2016 May 1;17(5):402-409. [doi: 10.1016/j.jamda.2015.12.013] [Medline: 26825684]
8. Cifrek M, Medved V, Tonković S, Ostojić S. Surface EMG based muscle fatigue evaluation in biomechanics. Clin Biomech (Bristol) 2009 May;24(4):327-340. [doi: 10.1016/j.clinbiomech.2009.01.010] [Medline: 19285766]
9. Skrzat JM, Carp SJ, Dai T, et al. Use of surface electromyography to measure muscle fatigue in patients in an acute care hospital. Phys Ther 2020 Jun 23;100(6):897-906. [doi: 10.1093/ptj/pzaa035] [Medline: 32157308]
10. Naeem J, Hamzaid NA, Islam M, Azman AW, Bijak M. Mechanomyography-based muscle fatigue detection during electrically elicited cycling in patients with spinal cord injury. Med Biol Eng Comput 2019 Jun;57(6):1199-1211. [doi: 10.1007/s11517-019-01949-4] [Medline: 30687901]
11. Sheng X, Ding X, Guo W, Hua L, Wang M, Zhu X. Toward an integrated multi-modal sEMG/MMG/NIRS sensing system for human–machine interface robust to muscular fatigue. IEEE Sens J 2021;21(3):3702-3712. [doi: 10.1109/JSEN.2020.3023742]
12. Otálora S, Segatto MEV, Monteiro ME, Múnera M, Díaz CAR, Cifuentes CA. Data-driven approach for upper limb fatigue estimation based on wearable sensors. Sensors (Basel) 2023 Nov 20;23(22):9291. [doi: 10.3390/s23229291] [Medline: 38005677]
13. Li N, Zhou R, Krishna B, et al. Non-invasive techniques for muscle fatigue monitoring: a comprehensive survey. ACM Comput Surv 2024 Oct 31;56(9):1-40. [doi: 10.1145/3648679]
14. Dingwell JB, Joubert JE, Diefenthaeler F, Trinity JD. Changes in muscle activity and kinematics of highly trained cyclists during fatigue. IEEE Trans Biomed Eng 2008 Nov;55(11):2666-2674. [doi: 10.1109/TBME.2008.2001130] [Medline: 18990638]
15. Marco G, Alberto B, Taian V. Surface EMG and muscle fatigue: multi-channel approaches to the study of myoelectric manifestations of muscle fatigue. Physiol Meas 2017 May;38(5):R27-R60. [doi: 10.1088/1361-6579/aa60b9] [Medline: 28199218]
16. Arntz A, Weber F, Handgraaf M, et al. Technologies in home-based digital rehabilitation: scoping review. JMIR Rehabil Assist Technol 2023 Jul 27;10:e43615. [doi: 10.2196/43615] [Medline: 37253381]
17. Ferdinand KC, Yadav K, Nasser SA, et al. Disparities in hypertension and cardiovascular disease in blacks: The critical role of medication adherence. J Clin Hypertens (Greenwich) 2017 Oct;19(10):1015-1024. [doi: 10.1111/jch.13089] [Medline: 28856834]
18. Liao Y, Thompson C, Peterson S, Mandrola J, Beg MS. The future of wearable technologies and remote monitoring in health care. Am Soc Clin Oncol Educ Book 2019 Jan;39(39):115-121. [doi: 10.1200/EDBK_238919] [Medline: 31099626]
19. Korada H, Maiya A, Rao SK, Hande M. Effectiveness of customized insoles on maximum plantar pressure in diabetic foot syndrome: A systematic review. Diabetes Metab Syndr 2020;14(5):1093-1099. [doi: 10.1016/j.dsx.2020.06.041] [Medline: 32652497]
20. Ostaszewski M, Pauk J. Estimation of ground reaction forces and joint moments on the basis on plantar pressure insoles and wearable sensors for joint angle measurement. Technol Health Care 2018;26(S2):605-612. [doi: 10.3233/THC-182507] [Medline: 29843283]
21. Valldecabres R, Richards J, De Benito AM. The effect of match fatigue in elite badminton players using plantar pressure measurements and the implications to injury mechanisms. Sports Biomech 2022 Sep;21(8):940-957. [doi: 10.1080/14763141.2020.1712469] [Medline: 32301399]
22. Haris F, Liau BY, Jan YK, et al. A review of the plantar pressure distribution effects from insole materials and at different walking speeds. Appl Sci (Basel) 2021;11(24):11851. [doi: 10.3390/app112411851]
23. Stolwijk NM, Duysens J, Louwerens JWK, Keijsers NLW. Plantar pressure changes after long-distance walking. Med Sci Sports Exerc 2010 Dec;42(12):2264-2272. [doi: 10.1249/MSS.0b013e3181e305f4] [Medline: 20404764]
24. Wang D, Cai P, Mao Z. The configuration of plantar pressure sensing cells for wearable measurement of COP coordinates. Biomed Eng Online 2016 Oct 26;15(1):116. [doi: 10.1186/s12938-016-0237-3] [Medline: 27784299]
25. Cruz-Jentoft AJ, Baeyens JP, Bauer JM, et al. Sarcopenia: European consensus on definition and diagnosis: Report of the European Working Group on Sarcopenia in Older People. Age Ageing 2010 Jul;39(4):412-423. [doi: 10.1093/ageing/afq034] [Medline: 20392703]
26. Padua DA, Arnold BL, Perrin DH, Gansneder BM, Carcia CR, Granata KP. Fatigue, vertical leg stiffness, and stiffness control strategies in males and females. J Athl Train 2006;41(3):294-304. [Medline: 17043698]

27. Merletti R, Farina D. Surface Electromyography: Physiology, Engineering, and Applications: John Wiley & Sons, Inc; 2004.
28. Barnett F. The effect of exercise on affective and self-efficacy responses in older and younger women. J Phys Act Health 2013 Jan;10(1):97-105. [doi: 10.1123/jpah.10.1.97] [Medline: 22396245]
29. Berglund A, Orädd H. Exploring the psychological effects and physical exertion of using different movement interactions in casual exergames that promote active microbreaks: quasi-experimental study. JMIR Serious Games 2024 Aug 26;12:e55905. [doi: 10.2196/55905] [Medline: 39186776]
30. Bisiaux M, Moretto P. The effects of fatigue on plantar pressure distribution in walking. Gait Posture 2008 Nov;28(4):693-698. [doi: 10.1016/j.gaitpost.2008.05.009] [Medline: 18586495]
31. Nagel A, Fernholz F, Kibele C, Rosenbaum D. Long distance running increases plantar pressures beneath the metatarsal heads: a barefoot walking investigation of 200 marathon runners. Gait Posture 2008 Jan;27(1):152-155. [doi: 10.1016/j.gaitpost.2006.12.012] [Medline: 17276688]
32. Sun D, Fekete G, Mei Q, Gu Y. The effect of walking speed on the foot inter-segment kinematics, ground reaction forces and lower limb joint moments. PeerJ 2018;6:e5517. [doi: 10.7717/peerj.5517] [Medline: 30155372]
33. Honert EC, Bastas G, Zelik KE. Effect of toe joint stiffness and toe shape on walking biomechanics. Bioinspir Biomim 2018 Oct 10;13(6):066007. [doi: 10.1088/1748-3190/aadf46] [Medline: 30187893]
34. Rangra P, Santos D, Coda A, Jagadamma K. The influence of walking speed and heel height on peak plantar pressure in the forefoot of healthy adults: a pilot study. Clin Res Foot Ankle 2017;05(2). [doi: 10.4172/2329-910X.1000239]
35. Zhang LY, Liu QL, Yick KL, Yip J, Ng SP. Analysis of diabetic foot deformation and plantar pressure distribution of women at different walking speeds. Int J Environ Res Public Health 2023 Feb 19;20(4):3688. [doi: 10.3390/ijerph20043688] [Medline: 36834384]
36. Mootanah R, Song J, Lenhoff MW, et al. Foot type biomechanics part 2: are structure and anthropometrics related to function? Gait Posture 2013 Mar;37(3):452-456. [doi: 10.1016/j.gaitpost.2012.09.008] [Medline: 23107624]
37. Korpelainen R, Orava S, Karpakka J, Siira P, Hulkko A. Risk factors for recurrent stress fractures in athletes. Am J Sports Med 2001;29(3):304-310. [doi: 10.1177/03635465010290030901] [Medline: 11394600]
38. Bercovitz T, Herman A, Solomonow-Avnon D, Wolf A, Kodesh E. Plantar pressure modifications in experienced runners following an exhaustive run. Sports Biomech 2022 Nov;21(10):1189-1199. [doi: 10.1080/14763141.2020.1743744] [Medline: 32321365]
39. Takahashi K, Demura SI, Aoki H. Effects of lower limbs exercise with light and heavy loads on the center of gravity sway. AJSSM 2021 Jan 15;9(1):8-12. [doi: 10.12691/ajssm-9-1-3]
40. Yoo JI, Kim H, Ha YC, Kwon HB, Koo KH. Osteosarcopenia in patients with hip fracture is related with high mortality. J Korean Med Sci 2018 Jan 22;33(4):e27. [doi: 10.3346/jkms.2018.33.e27] [Medline: 29318794]
41. Rogers A, Morrison SC, Gorst T, et al. Repeatability of plantar pressure assessment during barefoot walking in people with stroke. J Foot Ankle Res 2020 Jun 29;13(1):39. [doi: 10.1186/s13047-020-00407-x] [Medline: 32600388]
42. Maqbool B, Herold S. Potential effectiveness and efficiency issues in usability evaluation within digital health: A systematic literature review. J Syst Softw 2024 Feb;208:111881. [doi: 10.1016/j.jss.2023.111881]
43. Cirkel L, Lechner F, Schlicker N, et al. Adoption and perception of prescribable digital health applications (DiGA) and the advancing digitalization among German internal medicine physicians: a cross-sectional survey study. BMC Health Serv Res 2024 Nov 6;24(1):1353. [doi: 10.1186/s12913-024-11807-1] [Medline: 39506735]
44. Rochester L, Mazzà C, Mueller A, et al. A roadmap to inform development, validation and approval of digital mobility outcomes: the Mobilise-D approach. Digit Biomark 2020;4(Suppl 1):13-27. [doi: 10.1159/000512513] [Medline: 33442578]
45. Eken MM, Braendvik SM, Bardal EM, Houdijk H, Dallmeijer AJ, Roeleveld K. Lower limb muscle fatigue during walking in children with cerebral palsy. Dev Med Child Neurol 2019 Feb;61(2):212-218. [doi: 10.1111/dmcn.14002] [Medline: 30156008]
46. Liu SH, Lin CB, Chen Y, Chen W, Huang TS, Hsu CY. An EMG patch for the real-time monitoring of muscle-fatigue conditions during exercise. Sensors (Basel) 2019 Jul 14;19(14):3108. [doi: 10.3390/s19143108] [Medline: 31337107]
47. Sarker P, Mirka GA. The effects of repetitive bouts of a fatiguing exertion (with breaks) on the slope of EMG measures of localized muscle fatigue. J Electromyogr Kinesiol 2020 Apr;51:102382. [doi: 10.1016/j.jelekin.2019.102382] [Medline: 31981942]
48. Kallenberg LAC, Schulte E, Disselhorst-Klug C, Hermens HJ. Myoelectric manifestations of fatigue at low contraction levels in subjects with and without chronic pain. J Electromyogr Kinesiol 2007 Jun;17(3):264-274. [doi: 10.1016/j.jelekin.2006.04.004] [Medline: 16790358]
49. Enoka RM, Duchateau J. Muscle fatigue: what, why and how it influences muscle function. J Physiol 2008 Jan 1;586(1):11-23. [doi: 10.1113/jphysiol.2007.139477] [Medline: 17702815]
50. Del Vecchio A, Negro F, Felici F, Farina D. Associations between motor unit action potential parameters and surface EMG features. J Appl Physiol (1985) 2017 Oct 1;123(4):835-843. [doi: 10.1152/japplphysiol.00482.2017] [Medline: 28751374]

Abbreviations

EMG: electromyography
HA: hallux
IMU: inertial measurement unit
LT: lesser toe
M1: first metatarsal
M5: fifth metatarsal
MDF: median frequency
MH: medial heel
MMG: mechanomyography
NIRS: near-infrared spectroscopy
RPE: rating of perceived exertion
sEMG: surface electromyography

Edited by A Kushniruk; submitted 20.08.24; peer-reviewed by A Panahi, H Sozen; revised version received 21.11.24; accepted 23.11.24; published 07.01.25.

Please cite as:
Chen FY, Lin TY, Huang YC, Widianawati E
Effectiveness of Using a Digital Wearable Plantar Pressure Device to Detect Muscle Fatigue: Within-Subject, Repeated Measures Experimental Design
JMIR Hum Factors 2025;12:e65578
URL: https://humanfactors.jmir.org/2025/1/e65578
doi:10.2196/65578

Original Paper

Usability Testing of a Bystander Bullying Intervention for Rural Middle Schools: Mixed Methods Study

Aida Midgett[1,2*], EdD; Diana M Doumas[1,2*], PhD; Claudia Peralta[1], PhD; Matt Peck[3], PhD; Blaine Reilly[1], PhD; Mary K Buller[4], MS

[1]Boise State University, Boise, ID, United States

[2]Institute for the Study of Behavioral Health and Addiction, Boise, ID, United States

[3]University of Arkansas, Fayetteville, AR, United States

[4]Klein and Buendel, Golden, CO, United States

[*]these authors contributed equally

Corresponding Author:
Aida Midgett, EdD
Boise State University
301 University Drive
Boise, ID, 83706
United States
Phone: 1 426 1214
Email: AidaMidgett@boisestate.edu

Abstract

Background: Targets of bullying are at high risk of negative socioemotional outcomes. Bullying programming in rural schools is important as bullying is more prevalent in those schools compared to urban schools. Comprehensive, school-wide bullying programs require resources that create significant barriers to implementation for rural schools. Because technology-based programs can reduce implementation barriers, the development of a technology-based program increases access to bullying prevention in rural settings.

Objective: We aimed to conduct usability testing of a bystander bullying intervention (STAC-T). We assessed usability and acceptability of the STAC-T application and differences in usability between school personnel and students. We were also interested in qualitative feedback about usability, program features, and feasibility.

Methods: A sample of 21 participants (n=10, 48% school personnel; n=11, 52% students) recruited from 2 rural middle schools in 2 states completed usability testing and a qualitative interview. We used descriptive statistics and 2-tailed independent-sample *t* tests to assess usability and program satisfaction. We used consensual qualitative research as a framework to extract themes about usefulness, relevance, needs, barriers, and feedback for intervention development.

Results: Usability testing indicated that the application was easy to use, acceptable, and feasible. School personnel (mean score 96.0, SD 3.9) and students (mean score 88.6, SD 9.5) rated the application well above the standard cutoff score for above-average usability (68.0). School personnel (mean score 6.10, SD 0.32) and students (mean score 6.09, SD 0.30) gave the application high user-friendliness ratings (0-7 scale; 7 indicates highest user-friendliness). All 10 school personnel stated they would recommend the program to others, and 90% (9/10) rated the program with 4 or 5 stars. Among students, 91% (10/11) stated they would recommend the program to others, and 100% (11/11) rated the program with 4 or 5 stars. There were no statistically significant differences in ratings between school personnel and students. Qualitative data revealed school personnel and students found the application useful, relevant, and appropriate while providing feedback about the importance of text narration and the need for teacher and parent training to accompany the student program. The data showed that school personnel and students found a tracker to report different types of bullying witnessed and strategies used to intervene by students a useful addition to STAC-T. School personnel reported perceiving the program to be practical and very likely to be adopted by schools, with time, cost, and accessibility being potential barriers. Overall, findings suggest that the STAC-T application has the potential to increase access to bullying prevention for students in rural communities.

Conclusions: The results demonstrate high usability and acceptability of STAC-T and provide support for implementing a full-scale randomized controlled trial to test the efficacy of the application.

(JMIR Hum Factors 2025;12:e67962) doi:10.2196/67962

KEYWORDS

technology-based bullying intervention; STAC-T; usability testing; middle school; rural

Introduction

Background

National statistics indicate that bullying is a national public health issue in the United States, with 19.2% of students aged 12 to 18 years reporting being bullied at school in the past year [1]. Bullying peaks in middle school, with 26.5% of sixth-grade students reporting being a target of school bullying, followed by 26.3% of seventh graders and 25.1% of eighth graders. Among students who report being bullied, 21.6% report being bullied online. Findings from a meta-analysis examining consequences of bullying have indicated that students who are targets of bullying report a wide range of negative mental health outcomes, including symptoms of anxiety, posttraumatic stress, depressive symptoms, nonsuicidal self-injury, suicidal ideation, and suicide attempts [2]. Similarly, being a target of cyberbullying is associated with internalizing symptoms, suicidal ideation [3-5], and alexithymia and psychotic experiences [6]. Thus, it is imperative to develop effective interventions for middle school students that are accessible and easy to implement to reduce bullying and the associated negative consequences.

Youth in Rural Schools

Students attending school in rural communities are at high risk of experiencing both school bullying and cyberbullying [7-9]. According to US national statistics, the highest rates of bullying among rural youth in the past decade were reported in 2019, with 27.7% or rural students reporting being bullied compared to 22.4% of students in urban areas [10]. Although rates of bullying peaked in 2019 for both rural and urban students, the most recent US national statistics indicate that the prevalence of school bullying victimization continues to be higher among students in rural areas (23.8%) than among students in urban areas (19%) [1]. Furthermore, among targets of bullying, students attending rural schools are also more likely to report being bullied online (23%) compared to students attending urban schools (19.5%). Rural students also report a higher rate of being bullied with repetition (18.8%) compared to urban students (14.4%). Among middle school students attending schools in rural communities, bullying victimization is associated with poor school relationships, negative school experiences [11], and depression and anxiety [11,12]. These data suggest the importance of developing school-based bullying prevention programs specifically for students in rural communities.

School-Based Bullying Interventions

Comprehensive, school-wide interventions are effective in reducing bullying and the associated negative mental health outcomes [13]. Furthermore, bystander training (eg, teaching students who witness bullying to intervene in bullying situations) is an important intervention component [13]. Although up to 80% of students report witnessing bullying [14], only 20% intervene [15]. Because students report that they do not know how to intervene when they witness bullying [16], bystander training is a promising approach to bullying prevention. However, few comprehensive school-based programs incorporate bystander training. In addition, comprehensive, school-wide bullying prevention programs are expensive, complex, and time intensive and require extensive training [17]. Because these interventions require substantial resources, many schools face implementation barriers. Schools in rural communities may also face economic disparities, creating further implementation challenges [18], including a lower tax base, increased training costs due to bringing in expert trainers, frequent staff turnover, school closures, staff overload, and lack of program advocates in bullying prevention [19]. Challenges related to logistical problems, training requirements, and limited funding can negatively impact program adoption and sustainability [19].

Shifting to a Technology-Based Bullying Intervention

Technology-based interventions have the potential to improve access to programming and decrease implementation barriers experienced in rural communities [19]. Although some rural areas have higher rates of poor internet connectivity, eligible schools in rural communities can receive discounts for internet and broadband services [20]. Federal grants to build broadband infrastructure in rural areas are also available [21]. In addition, research conducted with key middle school personnel (ie, administrators, teachers, and school counselors) in rural communities indicates both a strong interest in technology-based bullying prevention programs and positive implementation conditions (eg, administrative support and technology readiness) [22]. Thus, most students in rural communities have access to the necessary infrastructure to support technology-based programs, and key personnel in rural middle schools indicate that schools are interested and ready for technology-based bullying interventions.

Although there is strong interest and need for online bullying prevention programming, very few US bullying prevention programs include a technology component. Programs that do offer it as an adjunct to in-person delivery [13]; often rely on simple texting, not multimedia interfaces; and they do not train student bystanders to intervene. For example, the Build Respect, Stop Bullying program for middle schools uses an online platform [23] but is part of a large program with staff or family components without bystander training. Other available technology programs include (1) an SMS text messaging program pairing youth with a “text buddy” [24]; (2) apps that encourage students to report cyberbullying, block websites attracting cyberbullying, and notify parents and school personnel of cyberbullying; and (3) online social media campaigns and educational resources (eg, videos, testimonies, and quizzes) [18,25,26]. Although there are programs that incorporate technology into bullying prevention and intervention, none appear to offer a route to a technology-based, interactive bystander training for middle school students.

The STAC Intervention

STAC [27] is a brief, stand-alone bystander intervention that includes didactic and experiential training followed by 2 booster sessions. The 75-minute didactic training includes education about bullying and cyberbullying, the consequences of bullying, and bystander roles and a description of the four STAC strategies: (1) "Stealing the show"—using humor or distraction to interrupt the bullying situation removing the attention away from the target, (2) "Turning it over"—informing an adult about the bullying and asking for help, (3) "Accompanying others"—befriending or providing support to the targeted student, and (4) "Coaching compassion"—gently confronting the perpetrator to increase empathy for the target. The experiential training comprises a series of role-plays during which students practice using the STAC strategies through bullying scenarios. The STAC training is followed by two 15-minute booster sessions to reinforce learning. The STAC intervention is effective in reducing bullying [28,29] as well as mental health risks for bystanders [30-34]. STAC has also been adapted to be culturally appropriate for middle school students in rural communities [35-37]. Research on the adapted STAC intervention demonstrates both bullying reduction [35,38] and improved mental health [35,39] among students trained in the program.

The Technology-Based STAC Intervention

The technology-based STAC intervention (STAC-T) is an online application developed to shift intervention delivery from in-person implementation to a technology-based format, thereby increasing accessibility and reducing barriers to intervention implementation. STAC-T is designed to be easily disseminated to large groups of students, who can access the intervention from a computer, tablet, or smartphone. In addition, the 40-minute STAC-T application is designed to be modular, increasing implementation flexibility. The initial training is followed by one 15-minute booster session designed to reinforce skill acquisition through virtual role-plays. The program is interactive, including knowledge checks, personalized feedback, and the selection of avatars to respond to bullying scenarios. Initial development included the design and testing of a STAC-T prototype. The design of STAC-T was developed based on the content of the in-person STAC intervention for rural middle schools as well as feedback from an expert advisory board and key middle school personnel in rural communities. In addition, students attending rural middle schools participated in 3 iterative focus groups, providing feedback on program usefulness, content, and functionality [40]. Once developed, the STAC-T prototype was evaluated through usability testing, which provided feedback from end users on program functioning [41]. The results from usability testing with key personnel and students from 2 rural middle schools indicated that the STAC-T prototype was easy to use, acceptable, and feasible, supporting the full-scale development of the STAC-T application [40].

This Study

Bullying is a significant public health concern for students attending rural schools [7-9]. Comprehensive bullying intervention programs that incorporate bystander interventions are the standard for practice [13]; however, they place a high demand on schools for implementation [17] and can contribute to disparities in rural schools [19-22]. STAC-T has the potential to reduce barriers and increase access to bullying prevention for middle school students in rural settings [40]. Therefore, the purpose of this study was to evaluate the usability and acceptability of the full-scale STAC-T application to determine readiness for a large, multisite randomized controlled trial to evaluate the efficacy of the STAC-T application for middle school students in rural communities. Usability testing is an important step in the process of intervention development as it predicts the likelihood of program adoption [42]. To achieve this aim, we implemented usability testing with key stakeholders (ie, school personnel and students) at 2 middle schools in rural communities in 2 states (N=21) using a mixed methods design. This study had the following objectives: (1) to assess usability and acceptability of the STAC-T application and (2) to assess differences in usability between school personnel and students.

Methods

Participants

Participants were key school personnel (ie, administrators, teachers, and school counselors; 10/21, 48%) and students (11/21, 52%) recruited from 2 middle schools in rural, low-income communities in the Northwestern and Southern regions of the United States. Between 5 and 10 usability testers are needed to identify most usability issues [43]. The schools were selected based on previous and ongoing research partnerships. The 2 schools were Title 1 schools, with 95% and 99% of the student population at the 2 schools eligible for reduced or free lunch. Among school personnel, the ages ranged from 26 to 55 years (mean 43.4, SD 9.8 years), and most (9/10, 90%) were female. School personnel self-reported ethnicity or racial background as White (5/10, 50%), Hispanic or Latino (3/10, 30%), and Black or African American (2/10, 20%). Among students, ages ranged from 11 to 15 years (mean 12.8, SD 1.3 years), with 36% (4/11) in grade 6, a total of 18% (2/11) in grade 7, and 45% (5/11) in grade 8. Students self-reported gender as female (6/11, 55%) and male (5/11, 45%). Students self-reported ethnicity or racial background as White (4/11, 36%), Black or African American (4/11, 36%), and Hispanic or Latino (3/11, 27%).

Development of the STAC-T Application

The translation from the STAC in-person intervention to STAC-T was guided by persuasive system design, a theoretical guide for translating clinical aims to health-related technology frameworks [44-46]. The STAC-T application was developed using Agile programming, a collaborative and incremental programming methodology [44-46]. The application was functional on all web browsers that support HTML5 and was built on a full-stack web application using HTML and JavaScript as the main interface. React.js was used as the front-end framework. The look and feel of the program were designed using Adobe Illustrator and Photoshop and developed using HTML elements plus SVG, PNG, JPG, WAV, MP4, and GIF images, audio, and video graphics. The system is accessible on desktop computers and iOS and Android tablets and smartphones. All design and programming elements were

aligned, and stakeholders' inputs were incorporated throughout the multistage development. Programmers produced the STAC-T application; alpha and beta tested it in-house for stability and code errors; tested it for usability; and revised it following an iterative, Agile production process.

In previous studies, as well as the in the iterative interviews and focus group conducted in this study, participants indicated that increasing program interactivity, adding more color, and including more realistic images such as avatars were important to increase engagement [40] and promote behavior change [47]. Therefore, design elements such as space (bright colors and visual space), components (realistic characters and familiar objects), and mechanics (actions reported by students that occur in rural middle schools) were established for the program features. In addition, STAC strategy practice was designed to require students to select an avatar, view bullying scenarios and select actions to operationalize the STAC strategy, view the avatar enacting the selected action, and receive feedback on its effectiveness. An artist hand illustrated and styled 6 avatars. The avatars had light-, medium-, and dark-colored hair in different styles as well as light, medium, and dark skin tones for students to choose from to best represent themselves and stimulate engagement. To reward learning and bolster adherence, "badges" (visual reward icons; eg, "Show Stealing Badge") were included as intermittent awards to encourage user engagement (Figure 1).

The STAC-T application content comprises three modules: (1) What is Bullying?—users are presented with background information on bullying, including bullying definitions (ie, physical, verbal, and relationship bullying as well as cyberbullying), bullying facts and statistics, characteristics of students who bully, and negative consequences of bullying; (2) What are Bystanders?—users are taught what a bystander is and how bystanders affect bullying outcomes (this module explains the 4 bystander roles: *assistants*, those who intentionally help the bully; *reinforcers*, those who are not directly involved in hurting another student but encourage the bully by standing around, laughing, or watching quietly; *outsiders*, those who do not take sides while witnessing bullying; and *defenders*, those who do something to stop the bullying situation or help the target in some way); and (3) STAC Strategies—users are introduced to the 4 STAC strategies, which are "Stealing the show," "Turning it over," "Accompanying others," and "Coaching compassion" (this module also includes STAC strategy practice using avatars selected by the user). The booster session includes additional practice with bullying scenarios and STAC strategy use.

Iterative interviews (15/21, 71% of the participants; 7/15, 47% female and 8/15, 53% male; 6/15, 40% White; 6/15, 40% Black or African American; 3/15, 20% Hispanic or Latino) and 2 rounds of iterative focus groups (20/21, 95%; 11/20, 55% female and 9/20, 45% male; 6/20, 30% White; 9/20, 45% Black or African American; 3/20, 15% Hispanic or Latino; 2/20, 10% other) conducted with middle school students attending schools in rural communities in 2 states informed program development before usability testing. Students participating in the interviews provided feedback on design aspects of the program, including color scheme, narration, and cartoons. Students were given a sample slide in 5 color schemes, 3 narrator voice samples, and 3 cartoon-style character depictions. Students were asked to rank a series of questions about each program aspect and then rank their preferences. Feedback and ranked choices were used to select color schemes, the narrator, and the program artist. Iterative focus groups were then conducted to gather feedback from students related to content and stylistic aspects of the program. Overall, the program was well received; students reported that the content was helpful and they liked the look and feel of the teacher who appears throughout the training. Students in the first round of focus groups provided specific feedback to incorporate more cyberbullying scenarios (eg, having bullies use their phones to record their peers without their knowledge), make the appearance of the characters more realistic (eg, changing clothing, adding eyes to all the characters, and changing hairstyles), and improve the function of the program to make navigation more user-friendly. Students also expressed disliking a particular activity, which was removed from the program. Students' feedback was incorporated into the program before conducting the second round of focus groups. The students in the second round of focus groups provided additional feedback about how to make the appearance of the characters more realistic (eg, adding emotion to the characters), as well as adding background images to make the scenarios look more like what they are used to seeing at school and school-related activities, such as sporting events (eg, adding teachers, lockers, wall hangings, and bulletin boards). They also provided specific feedback about how to make student behaviors more realistic (eg, having the target look sad instead of crying and changing the type of bullying from physical to verbal in front of adults). Input from the focus groups informed the development of the final STAC-T application used in this study.

Figure 1. Samples from the technology-based STAC application.

Procedures

Participant recruitment and usability testing occurred in 2024. The researchers provided the liaisons (eg, school counselor and principal) from each school with an email script describing the purpose and procedures of the study. Inclusion criteria for school personnel consisted of being employed as a principal, teacher, or school counselor at the partnering school; speaking English or Spanish; consenting to participate; and having a desire to make a positive difference in the school climate. For students, inclusion criteria consisted of being enrolled in the partnering school, speaking English or Spanish, having parental consent and student assent, and having a desire to make a positive difference in the school climate. To assess desire to make a positive difference, school liaisons were provided with rubrics developed by the research team to identify key school personnel and students who exhibited the following characteristics assessed by the rubric: caring for students; having a desire to be a positive influence on the school climate; being approachable to students; caring about addressing the problem of bullying; and having leadership qualities in the case of school personnel or leadership, maturity, responsibility, caring toward others, influence, and a desire to be a positive influence on their peers in the case of students. For each item, school personnel and students were assessed on a 3-point scale, which included the ratings of *yes*, *somewhat*, or *no* for each item described previously. School personnel and students who scored *yes* or *somewhat* on all inclusion criteria were eligible to participate. Exclusion criteria for the study for both school personnel and students included having participated in a previous STAC study, speaking a language other than English or Spanish, and not providing consent or assent to participate.

The school liaison used the inclusion and exclusion criteria and the rubric to identify and contact key school personnel and students and then used the script to invite them to participate in the study.

Similarly to our previous usability research [40], we conducted the usability testing and interviews remotely. Research supports remote usability testing as a viable approach to gather high-quality user experience feedback [48,49]. To mitigate technology-related problems, before the testing session, a team member worked with the school liaison to ensure that they could open and operate the program using the school's computers and firewall. When problems occurred during the testing session, participants switched to a different device or, in one case, rescheduled the testing session.

During the usability testing, participants were asked to review the entire STAC-T application, including the booster session. Participants were asked to talk aloud while completing the tasks, identifying problems and the solutions attempted. The researchers and users were on videoconference and shared their screens. The researchers could see what the participants were doing, and they were able to communicate with each other in real time. The researchers observed the users as they worked through the tasks and asked questions to gather more data. After completing the STAC-T application, participants were asked to complete a brief usability survey followed by a semistructured interview and then a demographic questionnaire. All participants were asked to provide information about their perceptions of (1) program utility, (2) relevance and appropriateness of program content, (3) ways in which they would improve the program, and (4) using a bullying and strategy use tracker after the training. School personnel were also asked about (1) their thoughts on implementation feasibility, (2) the likelihood of school program adoption, (3) their thoughts on companion trainings for teachers and parents, and (4) barriers to program use. All individual interviews lasted 1 hour and were video recorded.

Measures

Demographics

Participants self-reported their age, ethnicity or race, and gender. Students also reported their grade level.

Program Usability

Usability was assessed using the System Usability Scale (SUS) [50]. The SUS is a widely used 10-item validated tool that measures the usability and acceptability of technology-based programs. Responses are measured on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). To calculate the SUS score, responses are converted as follows: (1) for odd-numbered items, 1 is subtracted from the response; (2) for even-numbered items, 5 is subtracted from the response; (3) the converted responses are added; and (4) the total is multiplied by 2.5. The final SUS score ranges from 0 to 100. An SUS score of ≥68 is considered above average [51].

Program User-Friendliness

One item selected from previous usability research [40] was used to assess the user-friendliness of the program. Participants were asked to rate the user-friendliness through the following question—"Overall, I would rate the user-friendliness of this program as:"—using a 7-point scale ranging from 0 (*worst imaginable*) to 7 (*best imaginable*).

Program Satisfaction

In total, 2 items selected from previous usability research [40] were used to assess program satisfaction. Participants were asked the following question—"Would you tell your friends/colleagues to use the program?"—with *yes*, *no*, and *don't know* as response choices. Participants were also asked how many stars they would give the program (1 star being the lowest and 5 stars being the highest).

Interview Questions

Following usability testing sessions, participants were asked a series of open-ended questions about the utility and relevance of the application prototype, as well as ways to improve the application, likelihood of program use, and potential implementation barriers. School personnel and students were asked the following: (1) "Please talk about your perception of how useful this program could be to helping to address the problem of bullying at school," (2) "Please share your thoughts on whether you think the content of this program is relevant and appropriate for students at your school and your community," and (3) "Can you talk about ways that you would improve the program?" Students were also asked the following: "If your school asked you to continue using the tracker, would it be useful?" School personnel were also asked the following: (1) "What are your thoughts on how practical or workable you think it would be to use this program at your school?" (2) "What do you believe is the likelihood that your school would use this intervention?" (3) "Do you think an online, brief teacher training and parent training module would be a helpful addition to this program?" (4) "What, if anything, would keep you from using this program?" (5) "How would you envision using the tracker, if at all, after completing the training modules and the two boosters?"

Data Analysis

Quantitative Analysis

Quantitative data from the questionnaires were analyzed using SPSS (version 29.0; IBM Corp). Before conducting statistical analyses, we examined the data for missing data points. We found no missing data. The data were also examined for outliers, defined as >3.3 SDs above the mean [52]. We found no outliers. To ensure that continuous data met statistical assumptions for parametric statistical tests, we assessed acceptable normality by using established guidelines for examining skew and kurtosis [53]. All continuous variables were within the acceptable range for skew and kurtosis. Descriptive statistics were used and are presented separately for school personnel and students. We examined differences between school personnel and students using 2-tailed independent-sample *t* tests for continuous variables and chi-square analyses for categorical variables. All statistical assumptions were met for the *t* tests and chi-square analyses. We controlled for type I error using the Bonferroni correction. On the basis of the calculated Bonferroni correction, all analyses were considered significant at $P<.004$.

Qualitative Analysis

Qualitative data from open-ended questions were analyzed separately for school personnel and students. In total, 3 team members, 2 of whom conducted the usability tests, transcribed the data verbatim. We used consensual qualitative research as a framework for data analysis. We used thematic analysis [54,55] to identify, analyze, organize, describe, and report themes found within the qualitative data. A faculty member with expertise in qualitative data analysis, along with 2 graduate students, 1 PhD student and 1 masters of art in counseling student, with previous experience in qualitative data analysis, analyzed the data. The faculty member led the data analysis team. The team met 2 times via Zoom (Zoom Video Communications). During their first meeting, the faculty member discussed the analysis protocol with the 2 students, as well as expectations and biases that they needed to be aware of as they analyzed the data. Each team member analyzed the transcripts for the school personnel and students separately to arrive at initial themes for each open-ended question from the interview protocol. Next, the team met one more time via Zoom and conducted additional email communication over a 4-week period to arrive at a consensus on themes and frequency categories supported by participant quotations. During their meetings, the team members shared their themes for each question and discussed agreement or disagreement about themes. The analysts relied on participant quotes to resolve disagreements. Once the team members reached a consensus, an external auditor reviewed the interview transcripts and themes for the school personnel and students. The auditor agreed with the team's findings. Interview data were deidentified to ensure anonymity, and quotes were identified by participant type (ie, school personnel or students).

Ethical Considerations

This study was registered with ClinicalTrials.gov (NCT05572398). All research procedures were approved (101-SB21-205) by Boise State University's institutional review

board. Researchers obtained informed consent from school personnel and parental consent and student assent in the case of students. For students, school liaisons met briefly with potential participants identified through the inclusion criteria to explain the project, and interested students were sent home with a letter describing the project as well as an informed consent form for the parent or guardian to sign. English and Spanish translations were provided in schools with a large Hispanic student population. Parents were also emailed the information and consent form. Parents could choose to sign the consent form via pen and paper or electronically. If they did not provide consent electronically, students were asked to return the signed consent form to the school liaison. Parents were provided with the study principal investigator's contact information and were encouraged to contact her if they had any questions or concerns. Students who returned the signed consent form were then provided with an opportunity to assent immediately before the interview, focus group, or usability testing session. In addition, several steps were taken to protect confidentiality: participants were informed that they were free to refrain from answering any questions, all data were identified only by a personal identifier number, and all research team members completed required training in protection of human research participants. School personnel received a US $50 Amazon gift card as an incentive for participation in the usability testing and individual interview. There were no incentives for student participants.

Results

Quantitative Analysis

Program Usability

Usability scores on the SUS are presented in Table 1. Overall, scores for both school personnel and students suggested a very high level of usability, functionality, and acceptability. As shown in Table 1, there were no differences in the scores on any of the individual items or the SUS total score between school personnel and students, with both participant groups scoring the STAC-T application at a very high level of usability.

Table 1. Means and SDs of the System Usability Scale scores from school personnel and students[a].

	School personnel (n=10), mean (SD)	Students (n=11), mean (SD)	*t* test (*df*)	*P* value
"I think that I would like to use the program frequently."	4.60 (0.52)	4.18 (0.60)	–1.70 (19)	.11
"I found the program to be more complex than it needed to be."	1.20 (0.42)	1.64 (0.92)	1.37 (19)	.19
"I thought the program was easy to use."	4.90 (0.32)	4.91 (0.30)	0.07 (19)	.95
"I think that I would need the support of a technical person to be able to use this program."	1.10 (0.32)	1.36 (0.67)	1.13 (19)	.27
"I found the various functions in the program were well put together with each other."	4.40 (1.27)	4.55 (0.69)	0.33 (19)	.74
"I thought there was too much inconsistency in this program."	1.00 (0.00)	1.55 (1.04)	1.66 (19)	.11
"I imagine that most people would learn to use this program very quickly."	4.90 (0.32)	4.55 (0.82)	–1.28 (19)	.22
"I found the program very awkward to use."	1.00 (0.00)	1.18 (0.40)	1.42 (19)	.17
"I felt very sure that I could use the program correctly."	4.90 (0.32)	4.55 (0.69)	–1.49 (19)	.15
"I needed to learn a lot of things before I could get going with this program."	1.00 (0.00)	1.55 (0.93)	1.84 (19)	.08
System Usability Scale total score	96.00 (3.94)	88.64 (9.51)	–2.27 (19)	.04

[a]Responses were scored on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Program User Friendliness

School personnel and students rated the program highly on user-friendliness. Among school personnel, scores on user-friendliness ranged from 6.00 to 7.00 (mean 6.10, SD 0.32). Among students, scores on user-friendliness ranged from 6.00 to 7.00 (mean 6.09, SD 0.30). There were no differences in scores between school personnel and students on user-friendliness (t_{19}=–0.07; *P*=.95).

Program Satisfaction

Program satisfaction ratings are presented in Table 2. Overall ratings suggested that school personnel and students were satisfied with the program. There were no differences in scores between school personnel and students on program recommendation (χ^2_1=1.0; *P*=.33) or star ratings (χ^2_1=2.8; *P*=.25).

Table 2. Program satisfaction results for school personnel and students.

	School personnel (n=10), n (%)	Students (n=11), n (%)
Would recommend the program		
Yes	10 (100)	10 (91)
No	0 (0)	0 (0)
Unsure	0 (0)	1 (9)
Star rating		
1 star	0 (0)	0 (0)
2 stars	0 (0)	0 (0)
3 stars	1 (10)	0 (0)
4 stars	5 (50)	3 (27)
5 stars	4 (40)	8 (73)

Qualitative Analysis

Overview

Qualitative feedback for the STAC-T application supported the quantitative findings and was very positive overall, with both school personnel and students sharing the perception that the STAC-T application is useful, relevant, and appropriate, as well as providing feedback on ways to improve the program. In addition, school personnel shared positive thoughts about program practicality and adoption, including interest in a teacher and parent training, and discussed barriers to program adoption. Both school personnel and students talked about the benefits of tracking students' reports of different types of bullying and strategies they used to intervene both as part of the program and as a stand-alone feature to be used after the training. The results are presented in the following sections organized by the following themes: (1) usefulness, (2) relevance and appropriateness, (3) program improvement, (4) program tracker, (5) practicality and adoption, (6) teacher and parent training, and (7) barriers.

Usefulness

All school personnel (10/10, 100%) and students (11/11, 100%) indicated that the program was useful and increased students' knowledge to intervene in bullying situations. For example, a school personnel member shared the following:

> *...the program that you guys are creating, will definitely inform the students what they need to look for, how they can become an active positive person.*

A student also stated the following:

> *I think it'll be helpful by telling other kids that it's not right to bully other kids because you don't know how they feel, and you don't know what they go through.*

In terms of increasing knowledge, a school personnel member indicated the following:

> *I like how it has all the different ways for the students to see how you can step in you know...that there's things they can do.*

Another one added the following:

> *...it gives students the tools that they might feel like they lack in general when bullying happens.*

A student stated the following:

> *It can teach ways on how to and when bullying is happening, how do we handle it, and make it stop quicker.*

Another student added the following:

> *...if people are making fun of a person about how they look or the way they eat, and they post a video on social media, I can like easily screenshot, show it to the principal, my teachers, to get these people to stop and like get them to stop the bullying.*

Relevance and Appropriateness

All school personnel (10/10, 100%) and students (11/11, 100%) reported that the program content was relevant and relatable and taught students empathy and prosocial attitudes. For example, one school personnel member shared the following:

> *I think the content was really relevant. It's things that you actually see at school or that you hear about or that we, that get actually reported.*

Another school personnel member added the following:

> *Oh yeah, absolutely. I think that they [students] can relate to cafeteria situations, getting on the bus and you know posting things especially you know on social media or on Instagram or TikTok or whatever.*

A student shared the following:

> *I think that is really, really relevant and I think that a lot of the situations that were used in this app as examples can be used. They can be real life situations.*

In terms of empathy and prosocial attitudes, a school personnel member said the following:

> *I think this program would help...teach more empathy cause when students have more empathy, they're less likely to exhibit those bullying behaviors.*

A student indicated the following:

It basically says that bullying is not okay and if you do see it here are a few ways on how to stop it.

Program Improvement

Both school personnel (9/10, 90%) and students (6/11, 55%) offered feedback on how to improve the program and talked about the importance of having the program be fully narrated. School personnel talked about ways to improve student engagement and user experience. For example, a school personnel member stated the following:

...making it so that all of the pop-ups and scenes are narrated.

A student also said the following:

Once you add that narrative it is going to be good because most middle schoolers they aren't going to want to read it.

In terms of improving engagement and user experience, a school personnel member said the following:

Just there was one [activity] where you had to click the arrows to move on and I feel like just making it a little bit more simple.

Another school personnel member added the following:

Slowing down the captions on the cartoons because to allow kids to see the picture to get a frame of mind of what's going on and then read the words.

Program Tracker

All the school personnel (10/10, 100%) and students (11/11, 100%) indicated that they would find using a bullying and strategy use tracker useful if their school asked them to continue to use it. When asked about how they would envision using the tracker, school personnel indicated that they would use it for data collection for programmatic feedback and prevention as well as ongoing teaching and student support. For example, when asked about the usefulness of the tracker, a school personnel member stated the following:

Yeah, you know, I think that would be good.

Another one added the following:

You can bring that data up and say, you know, you know, we're, we're seeing this type of bullying going on.

A student said the following:

Oh, yeah. Because it would let more people know that bullying has been going on and stuff.

Practicality and Adoption

All the school personnel (10/10, 100%) agreed that the program would be practical and workable, and almost all school personnel (9/10, 90%) stated that they would be likely to use the program at their school. For example, one school personnel member said the following:

...our students would definitely get on and be able to, you know, go through that program without any problems at all.

Another school personnel member stated the following:

Practical. I think that it addresses the needs of our students in their day-to-day interactions.

In terms of likelihood of use, one school personnel member stated the following:

I think it would be highly likely.

Another school personnel member added the following:

This will be helpful and they [schools] will use it because it wouldn't take too much time away from the academics, academic goals that we have.

Teacher and Parent Training Modules

All school personnel (10/10, 100%) reported that a brief online teacher and parent training would be useful, and almost all (9/10, 90%) stated that it would provide a common language and a means for future collaboration. For example, one school personnel member shared the following:

It's teaching the parents and the teachers what that looks like, you know, conversation or tools to help our kids.

Another school personnel member added the following:

Yes. Yes, so that you have this so that we're all working together as a team. And using that common language.

Barriers

When asked about barriers to using the program, most school personnel members (7/10, 70%) reiterated that they would use the program, but many of them (8/10, 80%) discussed barriers. For example, one school personnel member stated the following:

I can't think of one negative reason or one reason I wouldn't want to use it.

However, other school personnel identified potential barriers, with one member stating the following:

Not being accessible on the devices that the kids have available.

Another school personnel member said the following:

Time. But I don't see that as a factor for us because we could fit it into our advisory class.

A third school personnel member added the following:

Only thing I can think of is funding.

Discussion

Principal Findings

The purpose of this study was to examine the usability of the STAC-T application, a technology-based bystander bullying intervention designed specifically for middle schools in rural communities. We were interested in feedback from middle school personnel as they are in the position of making decisions related to adopting and implementing bullying interventions and from students as end users. The primary aim of this study was to test the usability of the STAC-T application and assess

program utility, user-friendliness, and relevance as well as feasibility and ways to improve the program. Overall, both the quantitative survey results and qualitative interview findings indicate that participants perceived the STAC-T application to be useful, user-friendly, and appropriate for students at their schools and reported high levels of satisfaction with the program. The findings of this study indicate that the STAC-T application is relevant and feasible for implementation in middle schools in rural communities. The quantitative and qualitative findings are consistent with our previous usability testing research [40].

The findings of this study provide support for the usability of the STAC-T application. Both school personnel (mean 96.0, SD 3.9) and student (mean 88.6, SD 9.5) scores on the SUS demonstrated a very high level of usability, exceeding the standard cutoff score of 68 [51]. Both school personnel and students also rated the user-friendliness of the STAC-T application as very high, with all participants rating the program at ≥6 on a scale ranging from 0 to 7. We found no differences between school personnel and students on SUS scores or user-friendliness ratings, suggesting that both groups of users found the STAC-T application to be highly usable. The qualitative data supported these results, with both school personnel and students indicating that they perceived the program to be useful as well as relevant and appropriate for middle school students in rural communities. Furthermore, both school personnel and students reported high levels of satisfaction with the program, with 100% (10/10) of school personnel and 91% (10/11) of students indicating that they would recommend the program to others. Furthermore, most participants (20/21, 95%) gave the program 4 or 5 stars on a scale ranging from 1 to 5 stars. These findings are particularly important as usability and acceptability are associated with both program adoption and implementation [42].

In terms of practicality and adoption of the intervention, school personnel believed that their school would be likely to use the STAC-T application and identified cost, time, and access as potential barriers. Our results are aligned with those of previous studies that indicate that school administrators in rural communities feel favorably about adopting and implementing online programs to address the problem of bullying [20]. In addition, our findings are similar to those of previous studies that identify cost [7,20], time [56], and access to technology [57] as notable barriers to online programming implementation in schools located in rural areas.

School personnel and students reported positive perceptions of the STAC-T program, as well as feedback for program improvement. The results of the qualitative analyses showed that both school personnel and students thought that the STAC-T content was relevant and relatable and increased students' knowledge of how to intervene in bullying situations. School personnel also stated that the program taught students empathy and prosocial attitudes. Furthermore, school personnel stated that the program was practical and workable and would be used at their school and that teacher and parent trainings would be useful additions to STAC-T to provide a common language among stakeholders. These findings suggest a need for bystander bullying intervention programs in rural schools that teach students how to intervene when they witness bullying behaviors through conducting role-plays to practice strategies across different bullying scenarios. In terms of program improvement, both school personnel and students highlighted the importance of having the entire program narrated to students. School personnel also talked about ways to improve user engagement by simplifying and slowing down a few program activities. Both students and school personnel also saw value in the posttraining tracker. Including these program modifications is likely to increase user engagement [40], which, in turn, can influence positive behavior change [47], potentially increasing the efficacy of STAC-T.

Limitations

This study supports the usability, relevance, and feasibility of the STAC-T prototype. However, certain limitations must be noted. Participants were recruited from 2 schools in rural, low-income areas from 2 states, one in the Northwestern region and one in the Southern region of the United States. Although participants were recruited from 2 different states to increase generalizability, school personnel and students from different regions of the country may have a different perspective. Future research including a broader geographic sample would increase generalizability. Furthermore, most of the school personnel participants in this study were female, further limiting the generalizability of the study. In addition, because of the small sample size, we were unable to explore differences in demographic factors, limiting information that could guide improvements to the STAC-T application. Finally, it is also possible that social desirability influenced participants as they were aware that the goal of the study was to translate the in-person STAC intervention to a technology-based format.

Implications

This study has important implications for the development and implementation of STAC-T in middle schools in rural communities. First, participants provided very high usability ratings for the STAC-T application, with qualitative data supporting the usability, utility, and relevance of the program. The findings also provided valuable information about the program itself, including the need for program narration and the utility of a type of bullying and strategy tracker that could be used after the training. School personnel also provided feedback about the importance of both teacher and parent modules to foster collaboration. These modules could be developed and offered to schools as companion modules for STAC-T. School personnel could provide teachers with the STAC-T teacher training during professional development time at school and provide parents with a link to the parent training via email. Training teachers, parents, and students could provide a common language and increase buy-in from all stakeholders [58,59]. In addition, participants indicated that program implementation is feasible as long as the program is cost-effective, brief, and students can access it on their school devices. Translating STAC from an in-person modality to an online platform can help decrease barriers to implementation, increasing implementation feasibility for rural schools by decreasing program costs and reducing the demand on schools by decreasing the need for staff training, in-class time, and

expert support. Overall, the findings of this study provide valuable feedback and a strong scientific premise for moving forward with a large, multisite randomized controlled trial to examine the efficacy of the STAC-T application.

Conclusions

Bullying is a significant public health concern for students in rural middle schools. Training students to intervene and developing programming that increases access are important factors in addressing this problem. The findings of this study demonstrate the usability, relevance, and feasibility of the STAC-T application. The preliminary data from this study support conducting a large, multisite randomized controlled trial to assess the efficacy of the STAC-T application for middle school students in rural communities. Technology-based bullying interventions, and STAC-T in particular, could be an instrumental approach to decreasing educational and mental health disparities in rural schools and helping address the problem of bullying.

Acknowledgments

This research was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under award 2R42MD014943-02A1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

AM, DMD, and MKB are developers of the technology-based STAC program. All other authors declare no conflicts of interest.

References

1. Student reports of bullying: results from the 2022 school crime supplement to the national crime victimization survey. National Center for Education Statistics. URL: https://nces.ed.gov/pubs2024/2024109rev.pdf [accessed 2024-10-13]
2. Moore SE, Norman RE, Suetani S, Thomas HJ, Sly PD, Scott JG. Consequences of bullying victimization in childhood and adolescence: a systematic review and meta-analysis. World J Psychiatry 2017 Mar 22;7(1):60-76 [FREE Full text] [doi: 10.5498/wjp.v7.i1.60] [Medline: 28401049]
3. Fisher BW, Gardella JH, Teurbe-Tolon AR. Peer cybervictimization among adolescents and the associated internalizing and externalizing problems: a meta-analysis. J Youth Adolesc 2016 Sep 22;45(9):1727-1743. [doi: 10.1007/s10964-016-0541-z] [Medline: 27447707]
4. Kowalski RM, Giumetti GW, Schroeder AN, Lattanner MR. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. Psychol Bull 2014 Jul;140(4):1073-1137. [doi: 10.1037/a0035618] [Medline: 24512111]
5. Sticca F, Perren S. Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. J Youth Adolesc 2013 May 27;42(5):739-750. [doi: 10.1007/s10964-012-9867-3] [Medline: 23184483]
6. Movahedi N, Hosseinian S, Rezaeian H, Nooripour R. Mediating role of alexithymia in relationship between cyberbullying and psychotic experiences in adolescents. BMC Psychol 2024 Aug 31;12(1):465 [FREE Full text] [doi: 10.1186/s40359-024-01960-x] [Medline: 39217387]
7. Dulmus CN, Theriot MT, Sowers KM, Blackburn JA. Student reports of peer bullying victimization in a rural school. Stress Trauma Crisis 2004 Jan;7(1):1-16. [doi: 10.1080/15434610490281093]
8. Leadbeater BJ, Sukhawathanakul P, Smith A, Thompson RS, Gladstone EJ, Sklar N. Bullying and victimization in rural schools: risks, reasons, and responses. J Rural Community Dev 2013;8(1):31-47.
9. Nansel TR, Overpeck M, Pilla RS, Ruan WJ, Simons-Morton B, Scheidt P. Bullying behaviors among US youth: prevalence and association with psychosocial adjustment. JAMA 2001 Apr 25;285(16):2094-2100 [FREE Full text] [doi: 10.1001/jama.285.16.2094] [Medline: 11311098]
10. Burns E, Mann R, Yanez C. Student reports of bullying: results from the 2019 school crime supplement to the National Crime Victimization Survey. Institute of Education Sciences. URL: https://nces.ed.gov/pubs2022/2022031.pdf [accessed 2024-10-13]
11. Evans CB, Smokowski PR, Cotter KL. Cumulative bullying victimization: an investigation of the dose–response relationship between victimization and the associated mental health outcomes, social supports, and school experiences of rural adolescents. Child Youth Serv Rev 2014 Sep;44:256-264. [doi: 10.1016/j.childyouth.2014.06.021]
12. Smokowski PR, Cotter KL, Robertson C, Guo S. Demographic, psychological, and school environment correlates of bullying victimization and school hassles in rural youth. J Criminol 2013 Jun 11;2013:1-13. [doi: 10.1155/2013/137583]
13. Gaffney H, Ttofi MM, Farrington DP. Evaluating the effectiveness of school-bullying prevention programs: an updated meta-analytical review. Aggress Violent Behav 2019 Mar;45:111-133 [FREE Full text] [doi: 10.1016/j.avb.2018.07.001]
14. Wu WC, Luu S, Luh DL. Defending behaviors, bullying roles, and their associations with mental health in junior high school students: a population-based study. BMC Public Health 2016 Oct 10;16(1):1066 [FREE Full text] [doi: 10.1186/s12889-016-3721-6] [Medline: 27724881]

15. Waasdorp TE, Bradshaw CP. Examining variation in adolescent bystanders' responses to bullying. Sch Psychol Rev 2019 Dec 27;47(1):18-33. [doi: 10.17105/spr-2017-0081.v47-1]
16. Bauman S, Yoon J, Iurino C, Hackett L. Experiences of adolescent witnesses to peer victimization: the bystander effect. J Sch Psychol 2020 Jun;80:1-14. [doi: 10.1016/j.jsp.2020.03.002] [Medline: 32540087]
17. Reinke WM, Stormont M, Herman KC, Puri R, Goel N. Supporting children's mental health in schools: teacher perceptions of needs, roles, and barriers. Sch Psychol Q 2011 Mar;26(1):1-13. [doi: 10.1037/A0022714]
18. IT asset management for schools. NetSupport DNA. URL: http://www.netsupportdna.com/us/education/ [accessed 2021-09-21]
19. Salimi N, Karimi-Shahanjarin A, Rezapur-Shahkolai F, Hamzeh B, Roshanaei G, Babamiri M. Use of a mixed-methods approach to evaluate the implementation of violence and bullying prevention programs in schools. Educ Urban Soc 2020 Nov 11;53(6):607-628. [doi: 10.1177/0013124520972090]
20. Rachfal CL, Gilroy AA. Broadband internet access and the digital divide: federal assistance programs. Congressional Research Service. URL: https://sgp.fas.org/crs/misc/RL30719.pdf [accessed 2024-10-13]
21. USDA launches new program to create high-speed internet e-connectivity in rural America. U.S. Department of Agriculture. 2018. URL: https://www.usda.gov/media/press-releases/2018/12/13/usda-launches-new-program-create-high-speed-internet-e-connectivity [accessed 2021-09-21]
22. Midgett A, Doumas DM, Myers VH, Moody S, Doud A. Technology-based bullying intervention for rural schools: perspectives on needs, challenges, and design. Rural Ment Health 2021 Jan;45(1):14-30 [FREE Full text] [doi: 10.1037/rmh0000151] [Medline: 33777286]
23. Miller DT, Prentice DA. Changing norms to change behavior. Annu Rev Psychol 2016 Jan 04;67(1):339-361. [doi: 10.1146/annurev-psych-010814-015013] [Medline: 26253542]
24. Prescient and strategic intelligence marketing research web site. Digital Health Market. URL: https://www.psmarketresearch.com/market-analysis/digital-health-market [accessed 2021-08-20]
25. Fast facts: educational institutions. U.S. Department of Education, Institute of Education Sciences. URL: https://nces.ed.gov/fastfacts/display.asp?id=84 [accessed 2021-08-20]
26. Evers KE, Prochaska JO, Van Marter DF, Johnson JL, Prochaska JM. Transtheoretical-based bullying prevention effectiveness trials in middle schools and high schools. Educ Res 2008 May 28;49(4):397-414. [doi: 10.1080/00131880701717271]
27. Midgett A, Doumas D, Sears D, Lundquist A, Hausheer R. A bystander bullying psychoeducation program with middle school students: a preliminary report. Prof Couns 2015 Dec;5(4):486-500 [FREE Full text] [doi: 10.15241/AM.5.4.486]
28. Midgett A, Doumas DM, Johnston AD. Establishing school counselors as leaders in bullying curriculum delivery: evaluation of a brief, school-wide bystander intervention. Prof Sch Couns 2018 May 28;21(1):1-9. [doi: 10.1177/2156759X18778781]
29. Midgett A, Doumas DM, Trull R, Johnson J. Training students who occasionally bully to be peer advocates: is a bystander intervention effective in reducing bullying behavior? J Child Adolesc Couns 2017 Jan 19;3(1):1-13. [doi: 10.1080/23727810.2016.1277116]
30. Doumas DM, Midgett A, Watts AD. The impact of a brief, bullying bystander intervention on internalizing symptoms: is gender a moderator of intervention effects? Sch Psychol Int 2019 Feb 14;40(3):275-293. [doi: 10.1177/0143034319830149]
31. Midgett A, Doumas DM. The impact of a brief bullying bystander intervention on depressive symptoms. J Couns Dev 2019 Jun 13;97(3):270-280. [doi: 10.1002/jcad.12267]
32. Midgett A, Doumas DM, Trull R. Evaluation of a bystander bullying intervention program for elementary school students. Prof Sch Couns 2018 Feb 15;20(1):172-183. [doi: 10.5330/1096-2409-20.1.172]
33. Midgett A, Doumas DM, Trull R, Johnston A. A randomized controlled study evaluating a brief, bystander bullying intervention with junior high school students. J Sch Couns 2017;15(9):1-34 [FREE Full text]
34. Watts AD, Doumas DM, Midgett A. Efficacy of a brief, school‐based bystander bullying intervention on high school students' alcohol use. J Addict Offender Couns 2019 Oct 07;40(2):66-83. [doi: 10.1002/jaoc.12066]
35. Midgett A, Doumas DM. Acceptability and short-term outcomes of a brief, bystander intervention program to decrease bullying in an ethnically blended school in low-income community. Contemp Sch Psychol 2020 Sep 18;24(4):508-517. [doi: 10.1007/S40688-020-00321-W]
36. Midgett A, Doumas DM, Moran M, Gallo L. Evaluation of a brief, school-based bystander bullying intervention: a pilot study conducted at an ethnically-blended, low-income school. J Educ Psychol Res (Overl Park) 2020;2(2):52-60 [FREE Full text] [doi: 10.33140/jepr.02.02.01] [Medline: 38009113]
37. Moran M, Midgett A, Doumas DM, Moody S, Porchia S. A mixed method evaluation of a culturally adapted, brief, bullying bystander intervention for middle school students. J Child Adolesc Couns 2019;5(3):221-238 [FREE Full text] [doi: 10.1080/23727810.2019.1669372] [Medline: 33015352]
38. Moran M, Midgett A, Doumas DM. Evaluation of a brief, bystander bullying intervention (STAC) for ethnically blended middle schools in low-income communities. Prof Sch Couns 2020 Jul 20;23(1):2156759X2094064. [doi: 10.1177/2156759X20940641]
39. Midgett A, Doumas DM, Peralta C, Bond L, Flay B. Impact of a brief, bystander bullying prevention program on depressive symptoms and passive suicidal ideation: a program evaluation model for school personnel. J Prev Health Promot 2020 Jul 26;1(1):80-103 [FREE Full text] [doi: 10.1177/2632077020942959] [Medline: 33738442]

40. Doumas DM, Midgett A, Myers V, Buller MK. Usability of a technology-based bystander bullying intervention for middle school students in rural, low-income communities: mixed methods study. JMIR Form Res 2021 Oct 26;5(10):e32382 [FREE Full text] [doi: 10.2196/32382] [Medline: 34582355]
41. Wichansky AM. Usability testing in 2000 and beyond. Ergonomics 2000 Jul 10;43(7):998-1006. [doi: 10.1080/001401300409170] [Medline: 10929833]
42. Lyon AR, Bruns EJ. User-centered redesign of evidence-based psychosocial interventions to enhance implementation-hospitable soil or better seeds? JAMA Psychiatry 2019 Jan 01;76(1):3-4. [doi: 10.1001/jamapsychiatry.2018.3060] [Medline: 30427985]
43. Zichermann G, Cunningham C. Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps. New York, NY: O'Reilly Media, Inc; 2011.
44. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
45. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):28. [doi: 10.17705/1cais.02428]
46. Oinas-Kukkonen H, Harjumaa M. A systematic framework for designing and evaluating persuasive systems. In: Proceedings of the 3rd International Conference on Persuasive Technology. 2008 Presented at: PERSUASIVE '08; June 4-6, 2008; Oulu, Finland p. 164-176 URL: https://link.springer.com/book/10.1007/978-3-540-68504-3 [doi: 10.1007/978-3-540-68504-3_15]
47. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
48. Norman DA, Draper SW. User Centered System Design: New Perspectives on Human-Computer Interaction. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 1986.
49. Peterson C. Bringing ADDIE to life: instructional design at its best. J Educ Multimed Hypermedia 2003;12(3):227-241. [doi: 10.1007/978-0-387-09506-6_3]
50. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. London, UK: CRC Press; 1996.
51. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Usability Stud 2009;4(3):114-123 [FREE Full text]
52. Tabachnick BG, Fidell LS. Using Multivariate Statistics. 5th edition. Boston, MA: Allyn and Bacon; 2007.
53. Kline RB. Becoming a Behavioral Science Researcher, Second Edition: A Guide to Producing Research That Matters. New York, NY: The Guilford Press; 2009.
54. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
55. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
56. Beames JR, Johnston L, O'Dea B, Torok M, Christensen H, Boydell KM, et al. Factors that help and hinder the implementation of digital depression prevention programs: school-based cross-sectional study. J Med Internet Res 2021 Aug 27;23(8):e26223 [FREE Full text] [doi: 10.2196/26223] [Medline: 34448701]
57. Graves JM, Abshire DA, Amiri S, Mackelprang JL. Disparities in technology and broadband internet access across rurality: implications for health and education. Fam Community Health 2021;44(4):257-265. [doi: 10.1097/FCH.0000000000000306] [Medline: 34269696]
58. Midgett A, Doumas DM, Hausheer R. Development of a teacher module for a brief, bystander bullying intervention for middle schools: perspectives from school personnel. Contemp School Psychol 2022 Jun 23;27(3):491-504. [doi: 10.1007/S40688-022-00413-9]
59. Midgett A, Doumas DM, Hausheer R, Bond L, Buller MK, Peralta C, et al. Feasibility of a bullying bystander intervention (STAC) parent module for rural schools. J Prev Health Promot 2023 Apr 08;4(2):292-319 [FREE Full text] [doi: 10.1177/26320770231183961] [Medline: 38826496]

Abbreviations

SUS: System Usability Scale

Edited by A Kushniruk; submitted 24.10.24; peer-reviewed by A AL-Asadi, R Nooripour; comments to author 23.12.24; revised version received 13.01.25; accepted 18.01.25; published 21.02.25.

Please cite as:
Midgett A, Doumas DM, Peralta C, Peck M, Reilly B, Buller MK
Usability Testing of a Bystander Bullying Intervention for Rural Middle Schools: Mixed Methods Study
JMIR Hum Factors 2025;12:e67962
URL: https://humanfactors.jmir.org/2025/1/e67962
doi:10.2196/67962
PMID:

Original Paper

Evaluation of Satisfaction With a Secure, Connected Mobile App for Women in Assisted Reproductive Technology Programs: Prospective Observational Study

Pauline Plouvier[1], MD; Romaric Marcilly[2,3], PhD; Geoffroy Robin[1], MD; Chaymae Benamar[2,3], MErg; Camille Robin[1], MD; Virginie Simon[1], MD; Anne Sophie Piau[1]; Isabelle Cambay[1]; Jessica Schiro[2,3], PhD; Christine Decanter[1], MD

[1]Departments of Assisted Reproductive Technologies and Fertility Preservation, Jeanne de Flandre Hospital, Lille, France

[2]Inserm, CIC-IT 1403, F-59000 Lille, France

[3]Univ. Lille, CHU Lille, ULR 2694 - METRICS : Évaluation des technologies de santé et des pratiques médicales, F-59000 Lille, France

Corresponding Author:
Pauline Plouvier, MD
Departments of Assisted Reproductive Technologies and Fertility Preservation
Jeanne de Flandre Hospital
Avenue Eugène Avinée
Lille, F-59000
France
Phone: 33 0626394791
Email: pauline.plouvier@chu-lille.fr

Abstract

Background: Telemedicine has emerged rapidly as a novel and secure tool to deliver medical information and prescriptions. A secure, connected health care app (WiStim) has been developed in order to facilitate dialogue between patients and the medical team during an ovarian stimulation cycle for medically assisted reproduction (MAR).

Objective: This study aimed to evaluate the patients' and midwives' levels of satisfaction with the connected mobile app.

Methods: We conducted a prospective, observational, single-center study at Lille University Hospital, France. From May 1 to July 31, 2021, all women undergoing ovarian stimulation started to receive their treatment advice through the mobile app. A total of 184 women were included and they filled out the 30-item Usefulness Satisfaction and Ease-of-Use (USE) questionnaire, which examines the users' opinions in 4 dimensions: usefulness, ease of use, ease of learning, and satisfaction. The women also answered a series of closed and open questions. The 5 midwives in our assisted reproductive technology center filled out the French version of the 10-item System Usability Scale (SUS) when the app was implemented and then after 3 and 6 months of use. We also performed semistructured interviews with the midwives.

Results: Overall, 183 women using the app completed the questionnaire. None refused to use the app, and 1 withdrew from the study. The mean scores for the four USE dimensions were all significantly greater than 4, that is, the middle of the response scale. The women liked the app's ease of use, the access to tutorial videos, and the reminders about appointments and treatments. In particular, the women liked to be able to (re)read the information; this reassured them, might have reduced the number of missed appointments and treatments, and made them more independent during the day, especially when they were working. Some of the women regretted the loss of direct contact with the midwife. The mean SUS score was 76 (SD 13.54) at the start of the study, 75 (SD 17.16) after 3 months, and 84 (11.21) after 6 months. According to the adjective rating scale, these scores corresponded to good usability for the app. After the requisite training and a familiarization period, the midwives reported that using the app saved them 2 hours a day. The mobile app enabled better transmission of information and thus probably helped to decrease treatment errors.

Conclusions: The WiStim connected mobile app is one of the first reliable, secure apps in the field of MAR. The app reassured the patients during the ovarian stimulation. Women and the medical team considered that the app was easy and intuitive to use. Given the growth in demand for MAR programs and the medical team's workload, the time savings provided by the app constitute a nonnegligible advantage.

(JMIR Hum Factors 2025;12:e63570) doi:10.2196/63570

KEYWORDS

mobile apps; mHealth; mobile health; assisted reproductive technologies; evaluation; satisfaction; reproduction; fertility; ovarian stimulation; ease of use; usability; midwives; obstetrics; gynecology

Introduction

In 2018, the European Society of Human Reproduction and Embryology's 22nd report on medically assisted reproduction (MAR) in Europe highlighted a continuous increase in the number of treatment cycles and the broad range of techniques used [1].

Every day, around the world, thousands of women undergo hormone assays and ultrasound scans of the pelvis as part of their MAR program. These burdensome, complex procedures can generate stress and anxiety for the women and their partners [2-4]. The women consulting in MAR departments are young, active, and, in many cases, occupied by work and family activities. Furthermore, health care professionals have to deal with a growing administrative burden and increasing demand for MAR while still providing high-quality patient care and support.

A large number of "eHealth" smartphone apps are being developed [5-8]. WiStim is a secure mobile app created in 2016 to facilitate communication between MAR patients and medical teams. The app provides access to a variety of documents and media (eg, test results and tutorials on self-injection) and gives daily advice on treatment during the ovarian stimulation phase. The objectives are (1) secure communication with patients, (2) better traceability, (3) a lower frequency of treatment errors caused by poor understanding, and (4) time savings for the care team. At present, the patient has to pay a subscription fee; however, some hospitals are considering whether to subsidize or cover this fee.

During each in vitro fertilization, frozen embryo transfer, or intrauterine insemination cycle, the midwife calls the women daily in order to adjust the treatment, check on adherence, and answer any questions. These calls are time-consuming and can be replaced with a connected app.

The primary objective of the current study was to evaluate levels of satisfaction with the connected app, according to both the patients and health care professionals. The secondary objectives were to identify potential difficulties in the use of the app and to study any app-associated changes in the health care professionals' practices and work organization.

Methods

Overview

We conducted a prospective, observational, single-center study in the MAR department at Lille University Hospital (Lille, France) from May 1 to July 31, 2021.

Ethical Considerations

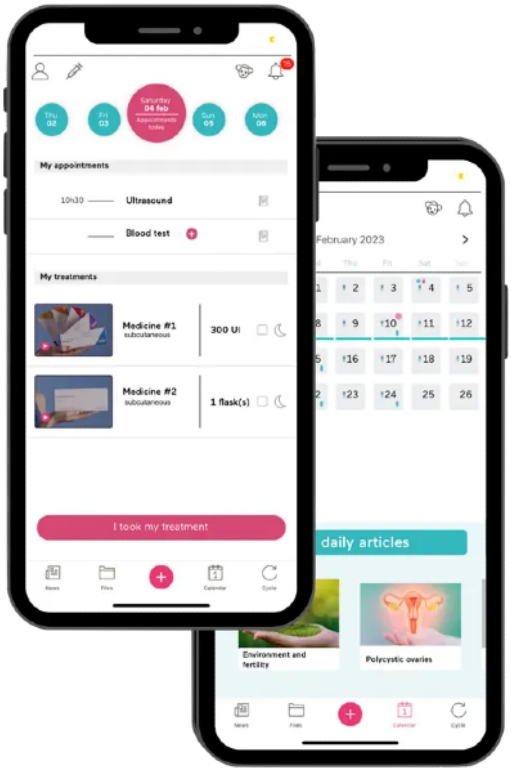
In line with the French legislation on human and social science surveys of routine clinical practice using anonymized personal data [9], approval by an independent ethics committee was neither required nor sought.

Connected Mobile App

The WiStim platform includes a mobile app for the patient (Figure 1) and a web-based management module for the medical team. The mobile app gives the patient access to advice on treatment, including the dose, the administration route, and any changes in the regimen. Tutorials on self-injections are also available. Every evening, the patient receives a reminder about her treatment and the date of her next appointment. The medical team sends information about the treatment and the next appointment through a secure web server (accredited by the French Ministry of Health) and is informed in real time about the woman's treatment.

The app requires a monthly subscription. However, Lille University Hospital has chosen to pay this cost on the patient's behalf during an initial test year.

Figure 1. The mobile app for the patient.



Patients

During a consultation, the gynecologist explained the app's principles to the patient and gave the latter an information sheet explaining how to download, install, and use the app. The patient then installed the app on her mobile phone and created an account. Once the stimulation phase had started, the patient received treatment advice through the app.

At the end of the MAR procedure, the women filled out a questionnaire (Multimedia Appendix 1). The questionnaire's closed questions concerned the type of MAR and the impact of using the app. The patient was able to write a comment for each closed question. Several open questions probed the woman's opinion of the app and the app's strengths and weaknesses. Next, the woman filled out the 30-item Usefulness Satisfaction and Ease-of-Use (USE) questionnaire, which measures the user's feelings in four dimensions: usefulness, ease of use, ease of learning, and satisfaction (Multimedia Appendix 2) [10,11]. The participants were not paid for their participation in the study.

Midwives

The five midwives in our MAR center filled out the French version of the System Usability Scale (SUS) when the app was implemented and then after 3 and 6 months of use [12]. The SUS (Multimedia Appendix 3) is a standardized, 10-item questionnaire for assessing the level of satisfaction with a technology's usability. For each of the 10 items, the midwives were invited to rate their level of agreement on a 5-point Likert scale ranging from "strongly disagree" to "strongly agree." We also performed semistructured interviews with the midwives 6 months after the implementation of the app in order to evaluate their opinion of the connected app, identify any changes in their work organization, and assess the app's acceptability (as defined in the unified theory of acceptance and use of technology [UTAUT]) [13]. The interviews were audio-recorded and transcribed.

Statistical Analyses

Statistical analyses were performed using Jamovi software (Jonathon Love, Damian Dropmann, and Ravi Selker, version 2.2.5, 2021), and thematic analyses of qualitative data were performed using QualCoder 3.0 (C Curtain). Data on the women's characteristics were used to describe the sample and perform subgroup analyses. The women's answers to the closed questions were analyzed using descriptive statistics, and the comments were analyzed qualitatively in order to highlight recurrent themes. Emerging themes in the qualitative analyses were not quantified; the goal was to understand the variety of the user's points of view. The data on the USE items and dimensions were analyzed with descriptive and inferential statistics. The threshold for statistical significance was set to $P<.05$. The results of the various questionnaires were compared in order to identify similarities and differences between the answers.

The SUS satisfaction score ranges from 1 to 100. Perceived satisfaction and usability are considered to be good when a score of 75 or more is obtained. To qualify the level of satisfaction, the mean SUS score was assessed against Aaron et al [14] grade scale and adjective rating scale. Semantic units (ie, sets of words representing the same idea) were extracted from the interviews with the midwives. An ergonomist attributed each semantic unit to one of the following themes: the opinion of the app, changes in practice, and dimensions of UTAUT acceptability (facilitating conditions, effort expectancy, social influence, and performance

expectancy). Within each theme, several subthemes were developed to represent the semantic units' diversity.

Results

Patients

Overview

During the recruitment period, 250 paper questionnaires were randomly distributed, and 184 patients responded (intrauterine insemination: 12%, in vitro fertilization: 73%, egg donation: 2%, frozen-thawed embryo transfer: 9%, fertility preservation: 4%). A total of 91 respondents had already been followed up by phone before the implementation of the app, 90 were monitored with the connected app alone, and 3 did not answer this item in the questionnaire. One woman discontinued use of the app prematurely, preferring direct contact with a midwife, and 2 women failed to complete the USE questionnaire.

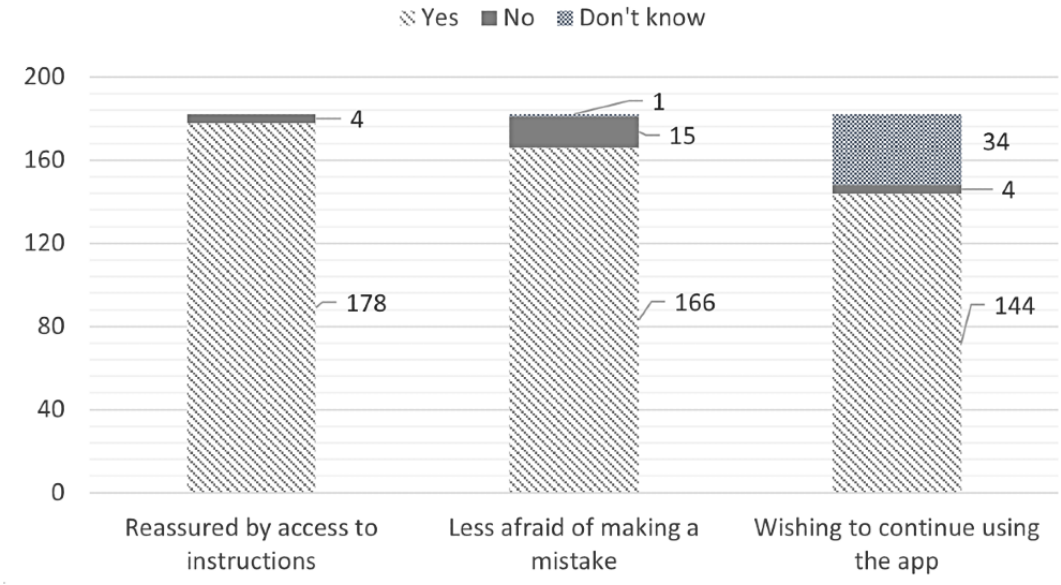
The mean scores for the four USE dimensions were all significantly greater than 4, that is, the middle of the response scale (usefulness, n=5.66, t_{180}=13.4, P<.001; ease of learning, n=6.25, t_{178}=18, P<.001; ease of use, n=5.96, t_{180}=16, P<.001; satisfaction, n=5.90, t_{177}=15.4, P<.001). The scores did not appear to be influenced by whether or not the women had been followed up by phone before the use of the app.

This good level of perceived usability was corroborated by our analysis of the women's comments and answers to the open questions. In all, 172 women described what they liked about the app. The app was seen as being easy to use in a guided, stepwise manner ("simple to use," "intuitive," and "clear"). Another reason for liking the app was its match with the women's needs and activities. The women mentioned the gain in efficiency linked to the app's centralization of information on treatment, follow-up, and appointments.

A total of 178 (97.8%) of the 182 respondents considered that the ability to access advice through the app was reassuring (Figure 2). The participants, having initially been followed up by phone, stated that they felt surer about the advice with the app because the instructions had to be rapidly written down during the phone call; with the app, the advice was available round the clock. The women also liked the additional information (eg, how to perform an injection) and the automatic reminders about appointments and treatments. A total of 166 (91.7%) of the 181 respondents considered that with the app, they were less stressed about making a treatment mistake. The participants liked the independence that the app provided; they no longer had to wait for the phone call from the midwife; previously, the call could come at any time in the afternoon.

Overall, 34 women did not reply to the question about wanting to continue to use the connected mobile app or not. 97.3% of the respondents said that they would be annoyed if they had to stop using the app (Figure 2). The 4 participants, who said they would feel relieved if they had to stop using the app, had all initially been followed up by phone and preferred direct communication with the medical team.

Figure 2. Answers to the following closed questions: "Does access to advice in the mobile app reassure you?" "Have you felt less stressed since you starting using the mobile app?" "Would you like to continue using the mobile app?".



Some weaknesses were mentioned: appointment errors, the lack of tutorials about some older models of self-injector pens, and late treatment reminders (ie, given after the scheduled time). Certain women expressed the need for direct interaction with the team of midwives via a chat function; the MAR department had decided not to offer this function because the women had already been given an email address for any questions.

A total of 53 (58.2%) of the 91 women who had initially been followed up by phone did not perceive any change in how they managed their treatment since the implementation of the app. In contrast, 38 (41.7%) women stated that the app modified the management of their treatment; they highlighted the lower mental burden, the easy, permanent access to information, the reassuring nature of the treatment reminders and app reminders,

and the fact their day no longer had to be organized around the call from the midwife.

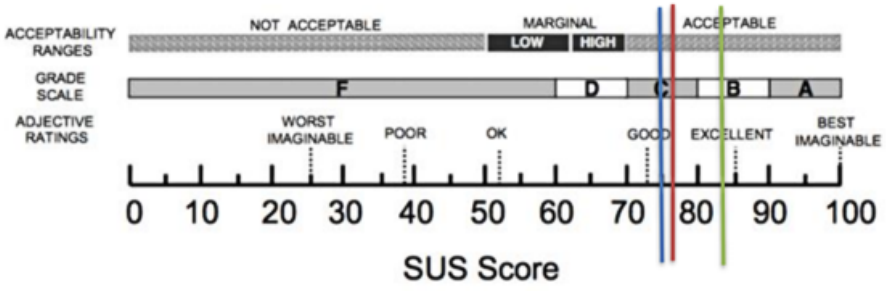
Midwives

The five midwives working in the MAR department at the time of the study filled out the SUS and were interviewed. The mean age was 43.6, and the midwives had been working in the department for an average of 6 years. The mean SUS score was 76 at the start of the study (n=4 respondents), 75 after 3 months (n=5), and 84 after 6 months (n=4; Figure 3). According to Aaron et al [14] adjective rating scale, these scores corresponded to good usability for the app.

A total of 207 semantic units were extracted from the semistructured interviews and analyzed. Even though the midwives stated that they needed some time to learn how to use the app and that the use of the app reduced the level of human contact with the women, they were generally very satisfied. The midwives found the interface pleasant to use and considered that the app was useful for the women and for the nurses who visited the women at home on a routine basis; the app modernized the midwives' practices.

During the interviews with the midwives, 3 of the 4 UTAUT dimensions emerged: facilitating conditions, effort expectancy, and performance expectancy. The social influence dimension did not emerge.

Figure 3. The mean System Usability Scale (SUS) score at the start of the study (76, SD 13.54) and after 3 months (75, SD 17.16) and 6 months (84, SD 11.21).



Facilitating Conditions

The midwives said that for the installation of the app, some IT equipment (computers) had been purchased and that they had attended training sessions on the use of the app. The midwives emphasized the good availability and responsiveness of the WiStim team when questions arose.

Effort Expectancy

The midwives viewed the app as a rather simple-to-use tool once they had been trained in its use. The app's different functions enable them to consult the patients' data, communicate with them, plan their appointments, and monitor their adherence to treatment.

Given that the app had not been integrated into the hospital information system at the time of the study, the midwives could not enter appointments into the hospital information system's diary from the app or vice versa; hence, to make appointments, they had to work on two IT tools in parallel. Even though the app had a simple-to-use interface, the fact that it was a web-based tool meant that the transitions were sometimes slow. Hence, data sometimes had to be entered twice: once on paper and then on the app.

Performance Expectancy

Use of the app was associated with a fall in the number of daily calls from midwives to patients. Nevertheless, the midwives continued to phone the women when the medical team had decided to stop the treatment or when the women had not ticked the boxes for adherence to treatment. This enabled them to optimize the dialogue by phone. Certain women continued to call or email the midwives regardless, to check that they had correctly understood the treatment procedures and advice specified by the app or to change an appointment. The midwives replied to these questions by email or by phone. The frequency of these requests tended to fall over time, as the midwives and the patients became more familiar with the app.

In order to ensure that the women were using the app correctly, the midwives helped some of them (especially those who did not understand or speak French sufficiently well) to install the app and explained how to use it. Furthermore, the advice stored in the app can be presented in several languages.

The decrease in the number of phone calls freed up an average of 2 hours per day, which enabled the midwives to perform other tasks (consultations, patient education sessions, etc).

Discussion

Principal Findings

On an annual basis, our center performs around 1400 oocyte retrieval procedures, 900 frozen embryo transfers, and 700 intrauterine inseminations. This corresponds to 30 to 50 women per day seen for ultrasound scans and hormone assays, and the volume of activity is increasing. Until now in France, MAR has only been available to heterosexual couples on medical indication. The law on bioethics, to be promulgated in 2021, extends MAR to female couples and single women. Activity in MAR centers has therefore risen sharply. Use of the connected mobile app was associated with a considerable decrease in the number of phone calls during the day. This resulted in significant time savings for the medical team, who were therefore able to perform other tasks. The women received their treatment advice in a secure manner, which had not been the case previously. The women felt reassured and less stressed about making treatment mistakes.

We did not question the partners, but in the open-ended questions, some of them said that they were reassured to be able to read the treatment instructions and felt more involved in the care compared with the phone call.

At present, it is difficult to quantify the extent of treatment errors during an ovarian stimulation program. Till date, no study has been published in the literature on this subject. Nevertheless, along with the treatment reminders, the fact that the woman's prescription is given in writing in the app and can be accessed at any time (and in several languages, if required) probably helped to reduce forgetfulness and treatment administration errors. This traceability is essential for secure communication between patients and professionals. Barrière et al [15] conducted an observational, real-life, longitudinal study involving 488 patients from 28 infertility centers in France to evaluate patient–infertility care provider relationships and communication in Assisted Reproductive Technology centers and investigate whether the quality of the care provided had an impact on patient adherence to treatment and monitoring protocols. They showed that even when patient-physician relationships appear to be satisfactory, patient miscomprehension and noncompliance during infertility treatment may be underestimated, and improvements in communication are also required.

Several studies [16,17] and reviews of the literature [18-20] have shown that stress is a major risk factor for impaired quality of life and a poor experience of infertility treatment. It also exposes patients to the risk of abandoning assisted reproduction procedures, thereby reducing their chances of pregnancy. It is therefore conceivable that mobile apps can reassure patients undergoing infertility treatments and consequently reduce the stress inherent in these procedures, as suggested in a recent review [21]. Further studies using standardized questionnaires and comparing the anxiety levels of the app users with those of nonusers could answer this hypothesis. The risk of dropping out of infertility management procedures could also be compared between these two groups.

Over the last decade, huge progress has been made in information and communication technology. "eHealth" and smart (connected) devices have brought together services that facilitate communication with the patient and thus improve the overall quality of care (teleconsultations, electronic health records, smart eHealth apps, etc) [22]. Telemedicine enables medical data to be shared with the patient or between health care professionals. Many connected health apps have been developed – notably in the fields of radiology and cardiology [23]. A recent study showed that connected health technologies can facilitate access to cancer care and improve the patient's psychological well-being and quality of life [24]. Recently, connected apps have been developed for use in gynecology and obstetrics [25]. There are probably other connected apps in the world in the field of fertility, but to date and to our knowledge, there are very few publications on this subject. Boivin et al [26] were interested in the development of a mobile app (MediEmo) to provide support during medically assisted reproduction. MediEmo is an app combining patient medication diary management and ease of integration into clinic systems with emotional support and data capture. They showed that 98% of patients expressed willingness to use the app, and almost 80% did so. Thus, the development of smartphone apps can contribute to fertility care and should be encouraged.

The profile of women in MAR programs is particularly suitable for this type of development: they are young and active and they all use a smartphone.

The protection of the patients' personal data is a crucial aspect in the development of connected health technologies [27,28]. WiStim complies with the European Union's General Data Protection Regulation, and the app's host is accredited by the French Ministry of Health.

One limitation is that the app is currently fee-based, although various funding options are being considered. In the present case, our hospital paid for the subscription to the app, and so it was cost-free for the users. However, the hospital's finance department is currently carrying out a medical-economic survey to assess the profitability of financing the app in relation to the time saved by the midwives. In fact, the midwife who used to call patients back can now do consultations instead.

Another limitation is the relatively small number of questionnaires analyzed and the fact that this was a single-center study. These are preliminary results, and they pave the way for further investigations in a larger population. Indeed, a multicenter study, with more questionnaires and over a longer period would be desirable to confirm these data.

Moreover, the questionnaires were distributed as soon as the app was installed in the department, so a learning curve was necessary for both patients and professionals. Overall satisfaction is probably higher now that the tool is better understood. Till now, all our patients have used the app.

Conclusion

Our results showed that the use of the connected health care app reassured women during the ovarian stimulation phase of a MAR program. Both the women and the medical team considered that the app was easy and intuitive to use. The information on treatment was sent in a secure manner and could be accessed around the clock by the patient. Connected health care apps have been developed widely in recent years. WiStim is one of the first reliable, secure apps in the field of MAR. Given the growth in demand for the MAR program and the medical team's workload, the time savings provided by the app constitute a nonnegligible advantage.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Patient questionnaire.
[DOCX File , 13 KB - humanfactors_v12i1e63570_app1.docx]

Multimedia Appendix 2
The Usefulness Satisfaction and Ease-of-Use questionnaire.
[DOCX File , 19 KB - humanfactors_v12i1e63570_app2.docx]

Multimedia Appendix 3
System usability scale.
[DOCX File , 15 KB - humanfactors_v12i1e63570_app3.docx]

References

1. European IVF Monitoring Consortium (EIM), for the European Society of Human ReproductionEmbryology (ESHRE), Wyns C, De Geyter C, Calhaz-Jorge C, Kupka MS, Motrenko T, et al. ART in Europe, 2018: results generated from European registries by ESHRE. Hum Reprod Open 2022;2022(3):hoac022 [FREE Full text] [doi: 10.1093/hropen/hoac022] [Medline: 35795850]
2. Stanhiser J, Steiner AZ. Psychosocial aspects of fertility and assisted reproductive technology. Obstet Gynecol Clin North Am 2018;45(3):563-574. [doi: 10.1016/j.ogc.2018.04.006] [Medline: 30092929]
3. Koumparou M, Bakas P, Pantos K, Economou M, Chrousos G. Stress management and In Vitro Fertilization (IVF): A pilot randomized controlled trial. Psychiatriki 2021;32(4):290-299 [FREE Full text] [doi: 10.22365/jpsych.2021.029] [Medline: 34390556]
4. Ying L, Wu LH, Loke AY. The effects of psychosocial interventions on the mental health, pregnancy rates, and marital function of infertile couples undergoing in vitro fertilization: a systematic review. J Assist Reprod Genet 2016;33(6):689-701 [FREE Full text] [doi: 10.1007/s10815-016-0690-8] [Medline: 26979745]
5. Meykiechel T, de Carne C, Gueguen I, Vatier C, Girard G, Buzzi JC, et al. Monitoring gestational diabetes mellitus patients with the telemedicine application MyDiabby decreases the rate of foetal macrosomia. Diabetes Metab 2022;48(2):101301. [doi: 10.1016/j.diabet.2021.101301] [Medline: 34757162]
6. Xie Q, Torous J, Goldberg SB. E-mental health for people with personality disorders: a systematic review. Curr Psychiatry Rep 2022;24(10):541-552 [FREE Full text] [doi: 10.1007/s11920-022-01360-1] [Medline: 35972718]
7. Karlsen IL, Svendsen PA, Abildgaard JS. A review of smartphone applications designed to improve occupational health, safety, and well-being at workplaces. BMC Public Health 2022;22(1):1520 [FREE Full text] [doi: 10.1186/s12889-022-13821-6] [Medline: 35948904]
8. Kumar N, Khunger M, Gupta A, Garg N. A content analysis of smartphone-based applications for hypertension management. J Am Soc Hypertens 2015;9(2):130-136. [doi: 10.1016/j.jash.2014.12.001] [Medline: 25660364]
9. Public Health Code: Section 1: Definitions. (Articles R1121-1 to R1121-2) [article in French] . Légifrance - Le service public de la diffusion du droit. URL: https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072665/LEGISCTA000006190943/#LEGISCTA000006190943
10. Lund AM. Measuring usability with the USE questionnaire. Usability Interface 2001;8(2):3-6 [FREE Full text]
11. Gao M, Kortum P, Oswald F. Psychometric evaluation of the USE (Usefulness, Satisfaction, and Ease of use) questionnaire for reliability and validity. Proc Hum Factors Ergon Soc Ann Mee 2018;62(1):1414-1418. [doi: 10.1177/1541931218621322]
12. Gronier GB, Baudet A. Psychometric evaluation of the F-SUS: creation and validation of the french version of the system usability scale. International Journal of Human–Computer Interaction 2021;37(16):1571-1582. [doi: 10.1080/10447318.2021.1898828]
13. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Quarterly 2003;27(3):425-478. [doi: 10.2307/30036540]
14. Aaron B, Kortum K, Miller M. Determining what individual SUS scores mean: adding an adjective rating scale. J. Usability Studies 2009;4(3):114-123 [FREE Full text]
15. Barrière P, Avril C, Benmahmoud-Zoubir A, Bénard N, Dejager S. Patient perceptions and understanding of treatment instructions for ovarian stimulation during infertility treatment. Reprod Biomed Soc Online 2019;9:37-47 [FREE Full text] [doi: 10.1016/j.rbms.2019.08.003] [Medline: 31993512]
16. Verberg MF, Eijkemans MJ, Heijnen EM, Broekmans FJ, de Klerk C, Fauser BC, et al. Why do couples drop-out from IVF treatment? A prospective cohort study. Hum Reprod 2008;23(9):2050-2055. [doi: 10.1093/humrep/den219] [Medline: 18544578]
17. Domar AD, Smith K, Conboy L, Iannone M, Alper M. A prospective investigation into the reasons why insured United States patients drop out of in vitro fertilization treatment. Fertil Steril 2010;94(4):1457-1459 [FREE Full text] [doi: 10.1016/j.fertnstert.2009.06.020] [Medline: 19591985]

18. Paraskevi L, Antigoni S, Kleanthi G. Stress and anxiety levels in couples who undergo fertility treatment: a review of systematic reviews. Mater Sociomed 2021;33(1):60-64 [FREE Full text] [doi: 10.5455/msm.2021.33.60-64] [Medline: 34012353]
19. Rich CW, Domar AD. Addressing the emotional barriers to access to reproductive care. Fertil Steril 2016;105(5):1124-1127 [FREE Full text] [doi: 10.1016/j.fertnstert.2016.02.017] [Medline: 27054306]
20. Cousineau TM, Domar AD. Psychological impact of infertility. Best Pract Res Clin Obstet Gynaecol 2007;21(2):293-308. [doi: 10.1016/j.bpobgyn.2006.12.003] [Medline: 17241818]
21. Meyers AJ, Domar AD. Research-supported mobile applications and internet-based technologies to mediate the psychological effects of infertility: a review. Reprod Biomed Online 2021;42(3):679-685. [doi: 10.1016/j.rbmo.2020.12.004] [Medline: 33487558]
22. Loncar-Turukalo T, Zdravevski E, Machado da Silva J, Chouvarda I, Trajkovik V. Literature on wearable technology for connected health: scoping review of research trends, advances, and barriers. J Med Internet Res 2019;21(9):e14017 [FREE Full text] [doi: 10.2196/14017] [Medline: 31489843]
23. Omboni S. Connected health in hypertension management. Front Cardiovasc Med 2019;6:76 [FREE Full text] [doi: 10.3389/fcvm.2019.00076] [Medline: 31263703]
24. Gitonga I, Desmond D, Duda N, Maguire R. Impact of connected health interventions on psychological wellbeing and quality of life in patients with cancer: a systematic review and meta-analysis. Psychooncology 2022;31(10):1621-1636 [FREE Full text] [doi: 10.1002/pon.6019] [Medline: 35996330]
25. DeNicola N, Marko K. Connected health and mobile apps in obstetrics and gynecology. Obstet Gynecol Clin North Am 2020;47(2):317-331. [doi: 10.1016/j.ogc.2020.02.008] [Medline: 32451020]
26. Robertson I, Harrison C, Ng KYB, Macklon N, Cheong Y, Boivin J. Development, implementation and initial feasibility testing of the MediEmo mobile application to provide support during medically assisted reproduction. Hum Reprod 2022;37(5):1007-1017 [FREE Full text] [doi: 10.1093/humrep/deac046] [Medline: 35348672]
27. Kamel Boulos MN, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform 2014;5(3):e61275. [doi: 10.5210/ojphi.v5i3.4814]
28. Schäfer F, Quinquis L, Klein M, Escutnaire J, Chavanel F, Chevallier H, et al. Attitudes and expectations of clinical research participants toward digital health and mobile dietary assessment tools: cross-sectional survey study. Front Digit Health 2022;4:794908 [FREE Full text] [doi: 10.3389/fdgth.2022.794908] [Medline: 35355684]

Abbreviations

MAR: medically assisted reproduction
SUS: System Usability Scale
USE: Usefulness Satisfaction and Ease-of-Use
UTAUT: unified theory of acceptance and use of technology

Edited by A Kushniruk; submitted 12.07.24; peer-reviewed by F Dénis, A Guerin, EM d'Argent; comments to author 19.10.24; revised version received 31.10.24; accepted 07.11.24; published 24.02.25.

Please cite as:
Plouvier P, Marcilly R, Robin G, Benamar C, Robin C, Simon V, Piau AS, Cambay I, Schiro J, Decanter C
Evaluation of Satisfaction With a Secure, Connected Mobile App for Women in Assisted Reproductive Technology Programs: Prospective Observational Study
JMIR Hum Factors 2025;12:e63570
URL: https://humanfactors.jmir.org/2025/1/e63570
doi:10.2196/63570
PMID:

Original Paper

Digital Cognitive Behavioral Therapy–Based Treatment for Insomnia, Nightmares, and Posttraumatic Stress Disorder Symptoms in Survivors of Wildfires: Pilot Randomized Feasibility Trial

Fadia Isaac[1], BA, BBS (Hons), MPsy; Britt Klein[1,2], PsyD; Huy Nguyen[2,3,4,5], PhD; Shaun Watson[1], PhD; Gerard A Kennedy[6], BBSc, PhD

[1]Institute of Health and Wellbeing, Federation University Australia, Victoria, Australia

[2]Health Innovation & Transformation Centre, Federation University Australia, Ballarat, Australia

[3]Department of Population and Quantitative Health Sciences, UMass Chan Medical School, Worcester, MI, United States

[4]Department of Epidemiology, Institute for Preventive Medicine and Public Health, Hanoi Medical University, Hanoi, Vietnam

[5]Faculty of Health Sciences, University of Cuu Long, Vinh Long Province, Vietnam

[6]School of Health and Biomedical Sciences, Royal Melbourne Institute of Technology University, Melbourne, Australia

Corresponding Author:
Fadia Isaac, BA, BBS (Hons), MPsy
Institute of Health and Wellbeing
Federation University Australia
University Drive
Mt Helen
Victoria, 3350
Australia
Phone: 61 353276717
Email: fadia.isaac@hotmail.com

Abstract

Background: Symptoms of insomnia, nightmares, and trauma are highly prevalent. However, there are significant barriers to accessing evidence-based treatments for these conditions, leading to poor mental health outcomes.

Objective: This pilot trial evaluated the feasibility of a 4-week, digital self-paced intervention combining cognitive behavioral therapy for insomnia and exposure, relaxation, and rescripting therapy for nightmares in survivors of wildfires from Australia, Canada, and the United States.

Methods: Study participants were recruited between May 2023 and December 2023 through social media platforms, workshops, conferences, and radio interviews. Participants had to meet at least one of the following criteria: a score of ≥8 on the Insomnia Severity Index, a score of ≥3 on the Nightmare Disorder Index, or a score of ≥31 on the PTSD Checklist for DSM-5. In total, 30 survivors of wildfires were allocated to either the treatment group (n=16, 53%) or the waitlist control group (n=14, 47%) in a sequential manner. Participants' ages ranged from 18 to 79 years, with a mean age of 52.50 (SD 16.26) years. The cohort consisted of 63% (19/30) female and 37% (11/30) male participants. Participants also completed self-report secondary outcome measures, including the Generalized Anxiety Disorder–7, the Patient Health Questionnaire–9, and the Pittsburgh Sleep Quality Index, via the HealthZone digital platform. Assessments were conducted at baseline, the posttreatment time point, and the 3-month follow-up, with the waitlist group undergoing an additional assessment at the pretreatment time point, after 4 weeks of waiting and before crossing over to treatment. This study used intention-to-treat analysis as a primary analysis and per-protocol analysis as a secondary analysis.

Results: Mixed-effects linear regression models and difference-in-differences analyses were used to assess the intervention's effects. The intention-to-treat analysis revealed significant improvements over time (main effect of time), with a 1.64-point reduction (P=.001) on the Nightmare Disorder Index and 10.64-point reduction (P=.009) on the PTSD Checklist for DSM-5 at the postintervention time point. No significant changes were observed in insomnia symptoms. On the secondary measures, there was an interaction effect of condition × time, with a 2.22-point reduction (P<.001) on the Pittsburgh Sleep Quality Index, and a main effect of time, with a 6.48-point reduction (P<.001) on the Patient Health Questionnaire–9. No changes were detected on

the Generalized Anxiety Disorder–7. The per-protocol analysis yielded comparable results for both the primary and secondary measures.

Conclusions: The findings of this pilot trial demonstrated a reduction in nightmares and trauma symptoms. Future research studies should aim at evaluating the intervention in a more definitive trial with a larger sample size.

Trial Registration: Australian New Zealand Clinical Trials Registry (ANZCTR) ACTRN12623000415606; https://anzctr.org.au/Trial/Registration/TrialReview.aspx?id=385054

(JMIR Hum Factors 2025;12:e65228) doi:10.2196/65228

KEYWORDS

insomnia; nightmares; posttraumatic stress disorder; PTSD; wildfires; cognitive behavioral therapy for insomnia; CBTi; exposure, relaxation, and rescripting therapy; ERRT; Sleep Best-i; mobile health; mHealth; digital health; computer; eHealth; bushfires

Introduction

Background

The consequences of wildfires can be devastating, resulting in significant losses of life, property, and livelihood. According to the Emergency Events Database, wildfires have been responsible for 1890 fatalities and 14,360 injuries between 2000 and 2023, with the highest tolls observed in Australia, New Zealand, Southern Europe, and North America [1]. The impact of wildfires extends far beyond the immediate loss of life and property, with survivors often experiencing a range of physical and emotional challenges that can have long-lasting effects. Physical injuries, such as bodily burns, smoke inhalation, and eye irritation, as well as increased risk of myocardial infarction and certain types of cancers, are a significant concern for survivors [2-4]. In addition, the proximity of wildfires to residential areas can have a profound impact on overall well-being, with research showing that life satisfaction decreases significantly in areas within 0 to 15 km of the fires [5]. The traumatic experience of a wildfire can also leave a lasting emotional impact, with survivors often reporting persistent feelings of fear and unsafety that can linger long after the initial event [5].

Beyond the physical and emotional challenges, many survivors face significant financial pressures, including the cost of rebuilding or relocating. For those who have lost their homes, the decision of whether to rebuild can be a difficult and emotional one, with financial pressures and the loss of a sense of security and community adding to the burden [3]. Others may experience stressors related to the loss of a workplace, hardship, and financial uncertainty. Therefore, the availability and timing of government relief and community support services are critical in facilitating the recovery of communities and individuals, with initiatives such as rehousing and recovery projects, psychological support, and primary producer repair and restoration [3,5] helping mitigate the impact of wildfires. However, when such support is not provided in a timely manner, emotional and psychological distress can intensify, leading to multiple psychological problems that can become highly resistant to treatment.

This is particularly concerning as research has shown that exposure to wildfires can have a significant impact on emotional well-being. Over the past decade, there has been a growing interest in investigating the impact of wildfire trauma on mental health [6,7]. Posttraumatic stress disorder (PTSD) is among the more severe mental health conditions that may arise immediately or in the months following exposure to traumatic events [7]. Exposure to wildfires, which pose a significant threat to life or well-being, can increase an individual's risk of developing PTSD. People with PTSD may exhibit symptoms such as reliving traumatic events through nightmares or flashbacks; sleeping difficulties; irritability; intrusive memories; feelings of horror, shame, and anger; avoidance of reminders of the traumatic event; alterations in arousal state or mood and cognition; and difficulties with work and daily activities [8].

Sleep disturbances are prominent features of PTSD and are frequently reported following trauma [9,10]. The onset of sleep disturbances following exposure to traumatic events is a strong predictor of the development of PTSD [11]. According to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*), insomnia and nightmares fall under the re-experiencing and the hyperarousal clusters of PTSD symptoms [8]. Insomnia is the most prevalent sleep disturbance in survivors of wildfires and is characterized by difficulty initiating or maintaining sleep or early-morning awakening [8]. Survivors of trauma also frequently experience nightmares, which are repeated distressing and well-remembered narrative dreams that lead to the individual waking up in a fight-flight state of alertness and having difficulty returning to sleep or even trying to avoid sleeping [8]. A study examining sleep disturbances and PTSD in survivors of wildfires revealed that approximately 49.2% (n=126) experienced symptoms of insomnia, 28.7% experienced nightmares, and 77.9% reported PTSD symptoms [12].

Studies suggest that cognitive behavioral therapy for insomnia (CBTi); cognitive behavioral therapy (CBT) for PTSD; and exposure, relaxation, and rescripting therapy (ERRT) for nightmares are effective treatments for insomnia, nightmares, and other trauma symptoms [13-18]. CBTi includes stimulus control, sleep hygiene, sleep restriction, cognitive restructuring, and relaxation training [19,20]. Nightmare treatment using ERRT is effective in giving those who experience nightmares a sense of control over the nightmare through writing benign dreams that can be rehearsed verbally or mentally [21]. The treatment focuses on addressing the themes presented in the nightmare and incorporates rescripting, imagery or verbal rehearsal, relaxation, and sleep hygiene elements [22,23].

Most research on trauma and sleep disorders has been conducted with military veterans, who have been the primary focus of clinical trials and studies in this area [24-26]. There is a need to focus on developing treatments for people who have experienced the trauma of natural disasters, including wildfires. There have only been 2 clinical trials addressing the treatment of sleep disturbances and PTSD in survivors of wildfires [27,28]. Both trials have shown that the administration of sleep dynamic therapy and CBTi therapist-assisted self-help interventions led to a significant reduction in insomnia and PTSD symptoms following treatment.

Objectives

While studies support the effectiveness of CBT, CBTi, and ERRT in treating trauma symptoms, insomnia, and nightmares, accessing these treatments can be challenging, particularly when wildfires cause destruction to infrastructure. This, in turn, causes difficulty in the delivery of counseling services to remote locations where large numbers of people are affected and in great need. Digital therapies can bridge this gap by delivering evidence-based treatments in communities affected by the trauma of wildfires [29,30]. However, the feasibility of digital therapies should be first examined in clinical trials before they are operationalized in the field [27,31]. Therefore, a brief, digital, self-paced, multicomponent therapeutic approach incorporating CBTi for insomnia, CBT for PTSD, and ERRT for nightmares was developed. The main objective of this pilot trial was to assess the feasibility of an intervention called Sleep Best-i in alleviating symptoms of insomnia, nightmares, and PTSD among survivors of wildfires. A secondary research objective was to undertake a comparative analysis of the degree of symptom reduction between the intervention and the waitlist control groups, with a view to informing an assessment of the feasibility of conducting a future definitive randomized controlled trial. Specifically, this assessment aimed to evaluate the acceptability and usability of the Sleep Best-i intervention. This clinical pilot trial tested the following hypotheses: (1) Sleep Best-i will result in significant reductions in insomnia, nightmares, and PTSD symptoms from baseline to the posttreatment time point compared to the waitlist group; (2) Sleep Best-i will also lead to significant reductions in anxiety and depression scores and improvements in sleep quality from baseline to the posttreatment time point compared to the waitlist group; and (3) both groups (within-group effects) will experience significant improvements in all measures (insomnia, nightmares, PTSD symptoms, anxiety, depression, and sleep quality) from baseline to the posttreatment time point, and these improvements will be maintained at the 3-month follow-up after receiving Sleep Best-i.

Methods

Study Design

This study used a parallel-arm, sequential alternation method of randomizing participants to either the intervention or waitlist group in a 50:50 ratio, with a crossover of the latter group. The alternation method ensures the ability to answer questions about the effectiveness of treatments to inform decision-making in clinical practice [32]. The treatment group completed self-report assessments at baseline, at the posttreatment time point following the administration of Sleep Best-i, and at the 3-month follow-up. The waitlist group completed the same assessments at baseline, at the pretreatment time point (following 4 weeks of waiting and before crossing over to treatment), at the posttreatment time point following the administration of Sleep Best-i, and at the 3-month follow-up. The CONSORT (Consolidated Standards of Reporting Trials) checklist was followed [33]. Please refer to Appendix 1 for the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (Multimedia Appendix 1).

Ethical Considerations

Following ethics approval from the Federation University Human Research Ethics Committee (approval 2022-153), an advertisement about the study was distributed. Survivors of wildfires interested in the study read a plain-language information statement (PLIS) about the study and provided baseline data. Participants who met the selection criteria provided informed consent by checking a box on a digital consent form hosted on Federation University's HealthZone platform, thereby indicating their agreement to take part in the study. Participants were informed that their participation was voluntary. To ensure the secure storage and handling of collected information, data were stored on a password-protected digital health platform (HealthZone) during the trial accessible only to the researchers involved in the study. To protect participants' confidentiality, names were replaced with unique identifiers, and only country of residence was collected, with no addresses recorded. The data will be retained for 15 years before being destroyed. To manage any potential risks, participants were provided with emergency contact numbers relevant to their country of residence. As a gesture of appreciation, participants who completed the study were offered an Aus $100 e-voucher (approximately US $66.70).

Participants

Participants were recruited between May 2023 and December 2023 via social media platforms, including international Facebook campaigns, radio broadcasts, LinkedIn, Reddit, and online community noticeboards. This study was also advertised on the Natural Hazards Research Australia and the Australian Institute for Disaster Resilience websites.

Adult survivors of wildfires from Australia, Canada, and the United States aged ≥18 years were invited to take part in the clinical pilot trial if they had experienced wildfires at any point in the past and had sleep or trauma symptoms.

An a priori power analysis indicated that a sample size of 50 was required to detect meaningful differences between the intervention and waitlist groups considering an α value of .05, a value of *d* of 1.0, an attrition rate of 20%, and 80% power [34].

Inclusion Criteria

To be eligible to take part in the trial, participants had to (1) provide digital consent; (2) have a score of ≥8 on the Insomnia Severity Index (ISI), (3) a score of ≥3 on the Nightmare Disorder

Index (NDI), or (4) a score of ≥31 on the PTSD Checklist for DSM-5 (PCL-5); (5) be fluent in the English language; and (6) have access to the internet. Exclusion criteria included not being a survivor of a wildfire, diagnosis of a psychotic disorder, suicide risk, diagnosis of sleep apnea or restless leg syndrome, diagnosis of alcohol or drug dependence, and attendance to psychotherapy for either sleep or PTSD conditions. In addition, participants were excluded if they were currently using steroids for any health condition; medications that promote sleep, such as benzodiazepines (eg, alprazolam, clonazepam, and temazepam); or medications that may affect sleep, such as opiates or other pain medications. By excluding these individuals, this study aimed to ensure that participants were suitable for the trial and that the results would not be confounded by other factors.

Procedure

The trial was registered prospectively in the Australian New Zealand Clinical Trials Registry (ACTRN12623000415606). Since its initiation, this study's protocol has undergone only minor revisions, specifically an update to the statistical analysis methodology to include linear mixed methods and intention-to-treat (ITT) analysis.

An advertisement of this pilot study with a URL was distributed on social media platforms displaying Federation University and the Natural Hazards Research Australia affiliations. This study was also promoted in national conferences, sleep workshops delivered to communities affected by wildfires, the Red Cross, and group meetings for people with sleep disorders in Australia. Radio interviews were also undertaken to promote the study. Survivors of wildfires who were interested read a PLIS (refer to Multimedia Appendix 2 for PLIS); signed a digital consent form (refer to Multimedia Appendix 3 for the consent form) to register an account on Federation University's digital HealthZone platform; and provided baseline data by completing a demographic questionnaire as well as self-report measures on insomnia, nightmares, and trauma symptoms. Participants had the option of providing a pseudonym if they wished. However, they needed to provide a valid email address to allow for communication about the study. Their email addresses were also directly connected to their personal dashboard on HealthZone. The baseline data were viewed by 2 researchers to decide eligibility. Those who were eligible (assessed by FI and GAK, both clinical psychologists) were contacted via email to notify them of their eligibility and their randomization and receive instructions on how to access their personal dashboard on HealthZone. Participants were sequentially allocated via simple randomization in the order of their enrollment (using a computer-generated simple randomization sequence), and they were informed about the purpose, allocation, and structure of the study as explained in the PLIS. However, they were not informed about the study's hypotheses. Once participants gained access to their personal dashboard within HealthZone, they were able to access information about the trial, complete the self-report secondary measures, and access the treatment modules consecutively. The treatment modules were provided at no cost, allowing participants to use their own devices and access the program and its modules on various platforms, including mobile devices, tablets, and desktop computers. HealthZone features a responsive web design compatible with both iOS and Android operating systems, making it easily accessible. Participant response data were captured on all devices. The modules were released sequentially to participants over 4 weeks. All data were collected through self-report measures at prespecified intervals. Participants were instructed to complete a module or a set of 2 modules each week along with the required interval assessments. Automated email reminders alerted participants about the release of each module and the specified assessments. If the modules were not viewed within the first 3 days of their release, an automated email reminder was sent to participants, and a second reminder was sent on the seventh day of their release. Automated email reminders at a similar frequency were also used for the primary and secondary measures (self-report scales). The personal dashboard was accessible to participants at any time. A "Contact us" tab was available on the personal dashboard that allowed participants to contact the research team and ask questions about the study or express any concerns. A total of 7% (2/30) of the participants were unsure about how to access the modules, and 3% (1/30) had an inquiry about one of the modules.

The research team sent 2 follow-up emails spaced 2 weeks apart to participants who registered, were randomized, but did not commence treatment, serving as a reminder about the study and to gauge their ongoing interest in participating.

The modules were available to participants from the start of the trial to the end of the follow-up period. Participants were able to access their dashboard for 4 weeks following the 3-month follow-up data collection. HealthZone recorded data about the number of log-ins, number of pages visited, and date of commencement and duration of participation for each participant. Each module was 17 minutes in duration. Therefore, participants were expected to spend at least 3 hours over the 4 weeks to indicate adherence to treatment. This time is optimal for viewing all modules and completing all assessments. No statistical analyses were conducted in the interim of the study. HealthZone was monitored daily during the trial to track any potential concerns or issues associated with the site or the participants.

Treatment Protocol

Sleep Best-i Intervention

Sleep Best-i was specifically designed for this clinical trial by some of the researchers who conducted this study (FI, BK, and GAK). The collated treatment manual draws from evidence-based treatment manuals authored by other sleep researchers [19,21,35-39]. The main therapeutic methods used in the manual were CBTi and ERRT. The recorded digital modules feature human-recorded voice and animated videos using VideoScribe (Sparkol) [40] that explain concepts related to sleep and trauma symptoms. The intervention also offers 2 role-plays of therapeutic sessions to demonstrate the administration of cognitive restructuring, dream rescripting, and sleep scheduling. Sleep Best-i consists of 6 modules administered over a 4-week period. In addition, the program is accessible via a platform that uses a responsive web design, with no requirement for a specific app or download. This allows participants to access the program from a variety of devices

(including mobile phones), as well as their preferred web browser. By being web based, the program can be easily accessed from anywhere at any time, making it a convenient and flexible intervention.

Participants were encouraged to complete the entire Sleep Best-i program, but completion of each module was not mandatory. This allowed individuals to engage with relevant content tailored to their specific needs, whether they experienced nightmares, PTSD, insomnia, or a combination of these conditions. Furthermore, participants had the flexibility to stop and start the program as desired, and all modules were available for the entire study period, enabling them to revisit specific modules should they wish to do so.

Table 1 shows the therapeutic modality offered in each module. Once released, the modules remained unchanged throughout the trial, and the intervention was successfully implemented without encountering technical difficulties. There were periodical checks to ensure that the modules were compliant with the treatment manual. All links have been archived in the Wayback Machine (Internet Archive).

Table 1. Sleep Best-i modules and the treatment offered in each module. All modules were approximately 17 minutes in duration.

Module	Strategies offered
Module 1 (psychoeducation)	Offered at week 1, the module provided psychoeducation about sleep and insomnia, the stages of sleep, and the neurobiology of sleep [14].
Module 2—part 1 (cognitive restructuring and sleep hygiene)	Offered at week 1, the module focused on the cognitive component of CBTi[a], types of unhelpful thoughts, methods to challenge unhelpful thoughts, and sleep hygiene [30].
Module 2—part 2 (sleep scheduling and stimulus control)	Offered at week 2, the module educated participants about specifying regular sleeping and waking up times with as little variation as possible between the 2. Stimulus control restricted the bedroom or the bed to sleep only [34].
Module 3 (trauma, PTSD[b], and flashbacks)	Offered at week 3, the module provided psychoeducation about trauma, PTSD, flashbacks, and how trauma leads to sleep difficulties. It also provided behavioral interventions for trauma symptoms.
Module 4 (nightmares)	Offered at week 4, the module explored nightmare disorder, how nightmares develop, and how to rescript the nightmare into a more neutral dream [16].
Module 5 (relapse prevention)	Offered at week 4, the module focused on identifying early warning signs and methods preventing relapse of symptoms [33].
Mindfulness module	This module offered a recorded progressive muscle relaxation mindfulness, and it was available to participants throughout the study.

[a]CBTi: cognitive behavioral therapy for insomnia.

[b]PTSD: posttraumatic stress disorder.

Waitlist Control Group

Following a waiting period of 4 weeks, the waitlist control group received Sleep Best-i in the same sequence in which it was released to the treatment group.

Measures

Overview

All scales were digital, self-report measures that were administered through HealthZone. The self-report measures as well as the modules were tested by 2 research team members (FI and BK) before being released to participants. The order of the modules and the questions were the same for all participants in the 2 groups.

Demographic Data

Participants provided the following information: name, email address, age, sex, employment status, educational level, marital status, country of residence, experience with wildfires, history of taking steroids, diagnosis of sleep apnea or restless leg syndrome, diagnosis of a psychotic disorder, use of alcohol or drug dependence, type of medications used to assist with sleeping, history of antidepressant use, and whether they were attending psychotherapy for sleep or PTSD.

Primary Measures

ISI Measure

The ISI [41] includes 7 items that assess insomnia severity through the subjective experience of sleep. The score varies from 0 to 28, with higher scores indicating more severe insomnia symptoms. A score of ≥8 indicates the presence of a subthreshold insomnia. Previous studies have demonstrated strong internal consistency for the ISI, with Cronbach α values ranging from 0.87 to 0.92, indicating high reliability [42].

NDI Measure

This scale consists of 5 items that assess symptoms of nightmares according to the DSM-5 criteria by screening for the occurrence of nightmares in a given week. The 5 items are summed to obtain a score between 0 and 20, with higher scores representing greater nightmare severity. The NDI exhibits good internal consistency, with a Cronbach α of 0.80 [43].

PCL-5 Measure

The PCL-5 [44] includes 20 self-report items assessing the presence of PTSD symptoms over a 1-month interval. A global score that ranges between 0 and 80 is obtained by summing the 20 items. A cutoff score of 31 provides optimal sensitivity and specificity in providing a provisional diagnosis of PTSD [45].

The PCL-5 has good psychometric properties, with an internal consistency of Cronbach α=0.94 [46].

Secondary Measures

Generalized Anxiety Disorder–7

This scale is a 7-item, self-report measure that assesses anxiety symptoms over a 2-week interval. A score from 0 to 21 is obtained by summing the 7 items. The Generalized Anxiety Disorder–7 (GAD-7) [47] has a good internal consistency, with a Cronbach α ranging between 0.82 and 0.93 [48]. A cutoff score of 10 discriminates between mild and severe symptoms of anxiety [48].

Patient Health Questionnaire–9

This scale consists of 9 items assessing symptoms of depression over the last 2 weeks. The overall score is obtained by summing all the 9 items and ranges from 0 to 27. A cutoff score of 10 is considered optimal in discriminating between mild and severe symptoms of depression [49]. In this study, the Cronbach α for the Patient Health Questionnaire–9 (PHQ-9) was 0.86.

Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) [50] consists of 19 items assessing sleep quality and disturbances related to sleep over a 1-month interval. A global score ranging between 0 and 21 is generated by summing 7 components of sleep quality. A global PSQI score of >5 is considered a sensitive and specific cutoff to discriminate between good and poor sleep quality [51]. The PSQI has a good internal validity, with the Cronbach α ranging between 0.79 and 0.81 [51].

Satisfaction and Level of Engagement With Treatment

One question was designed for this study to measure satisfaction with Sleep Best-i. Participants were asked to rate how likely they were to revisit the Sleep Best-i modules on a 5-point Likert scale from *strongly disagree* to *strongly agree* (0-4). To assess level of engagement with treatment, the number of log-ins and amount of time that each participant spent on the site were monitored.

Data Analysis

Data analyses were conducted using the Stata statistical software package (version 18; StataCorp) [52]. All variables were inspected for normality and outliers. An inspection of histograms and *Q-Q* plots revealed normally distributed data. Upon inspection of box plots on all scales, 1 outlier was identified on the PCL-5 at the 3-month follow-up assessment, a second outlier was detected on the PHQ-9 at the 3-month follow-up, and a third outlier was detected on the PSQI at the waitlist assessment (3-month follow-up assessment). The results did not differ upon removing the outliers from the analyses; therefore, the outliers were retained [53,54]. To ensure a robust evaluation of the intervention's feasibility, the ITT analysis with 30 participants and the per protocol (PP) analysis with a sample of 20 were used, allowing for a comprehensive assessment of treatment outcomes, thereby enhancing the accuracy and generalizability of the findings [33,55,56].

For both analyses, missing data were addressed using multiple imputation by chained equations. This approach is more robust than other methods such as single imputation because it generates multiple predictions for each missing value, accounting for uncertainty in the imputations and yielding more accurate SEs [57]. To impute missing data, background variables such as age, sex, educational level, employment status, country, and relationship status; baseline measures; and the most recent observations of each outcome were used.

To address uncertainty, each missing value was imputed 20 times using predictive mean matching based on the 3 nearest values across multiple iterations to create 100 imputed datasets. The imputed values were then averaged to create the final estimates [58]. Predictive mean matching, a partially parametric method, is preferable to fully parametric linear regression as it remains effective even when the normality assumption of the underlying variable is violated. This method also helps preserve the distribution of observed values in the missing data [59]. Missing value analysis indicated that the data were missing completely at random (Little missing completely at random test: $\chi^2_{867}=0.0$, $P>.99$), meaning that the differences between the missing and observed data were related to observed characteristics [60].

Adjusted analyses of intervention effects were conducted using mixed-effects linear regression models. Both fixed and random effects were estimated to assess the impact of the intervention on the change in all outcomes (primary and secondary) over time and by condition. Given that observed covariates were balanced between the 2 conditions (waitlist and treatment), the difference-in-differences (DID) effect for the fixed part of the mixed-effects models was tested by including interaction terms for time point and condition. DID is considered the most appropriate measure to assess causal effects in time-series designs as it effectively controls for time-invariant confounding variables. By doing so, DID reduces bias and enhances internal validity, providing a more accurate estimate of the treatment effect [61].

The appropriateness of the mixed-effects model for each outcome was assessed using chi-square statistics ($P<.05$). Random-intercept linear models were compared with random-intercept, random-slope models using likelihood ratio tests. The random-intercept, random-slope model was reported only if the likelihood ratio test indicated $P<.05$; otherwise, the simpler random-intercept model was selected. A significant effect of the intervention or difference in outcome between groups was assessed using a 2-tailed test of significance with $P<.05$.

A sensitivity analysis was conducted to test whether the model and covariance structures were misspecified. Several major types of covariance structures, including independent, unstructured, exchangeable, identity, and autoregressive structures, were examined. In addition, the model parameters using both maximum likelihood and restricted maximum likelihood methods were estimated. Although the results were generally consistent across different covariance structures, the model with an exchangeable structure was deemed the best fit.

Clinical Significance

Clinical significance was assessed for the 2 groups from baseline to the 3-month follow-up. To evaluate the clinical significance on sleep and trauma measures, specific criteria were used for each scale. For example, for the ISI, a cutoff score of ≤8 with a ≥6-point reduction in scores was used to assess whether participants reached a minimal clinically significant change (MCSC) [26,62]. For the NDI, there are no established criteria for MCSC; therefore, a cutoff score of ≤7 as suggested by Dietch et al [43] was used. Marx et al [63] recommend a score of ≤28 and a drop of ≥18 points on the PCL-5 as a guide for MCSC.

Results

Overview

A total of 33 participants responded to the advertisement, registered an account with Federation University's digital HealthZone platform, and provided baseline data. In total, 9% (3/33) were excluded because they did not meet the selection criteria, and 91% (30/33) were randomized to either the intervention or the waitlist group. A total of 53% (16/30) were randomized into the intervention group, and 47% (14/30) were in the waitlist group (for the participant CONSORT chart, refer to Figure 1). Ages ranged between 18 and 79 years (mean 52.50, SD 16.26 years; N=30), with most participants (19/30, 63%) being female (vs 11/30, 37% male). A total of 67% (20/30) of the participants completed the trial, and their data were analyzed as PP. The ages of the latter group ranged from 18 to 79 years (mean 53.75, SD 16.54 years), with most participants (14/20, 70%) being female (6/20, 30% were male). Demographic variables for the treatment and waitlist groups are shown in Table 2.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart showing participant flow through allocation to the treatment and waitlist conditions. ISI: Insomnia Severity Index; ITT: intention to treat; NDI: Nightmare Disorder Index; PCL-5: PTSD Checklist for DSM-5; PP: per protocol.

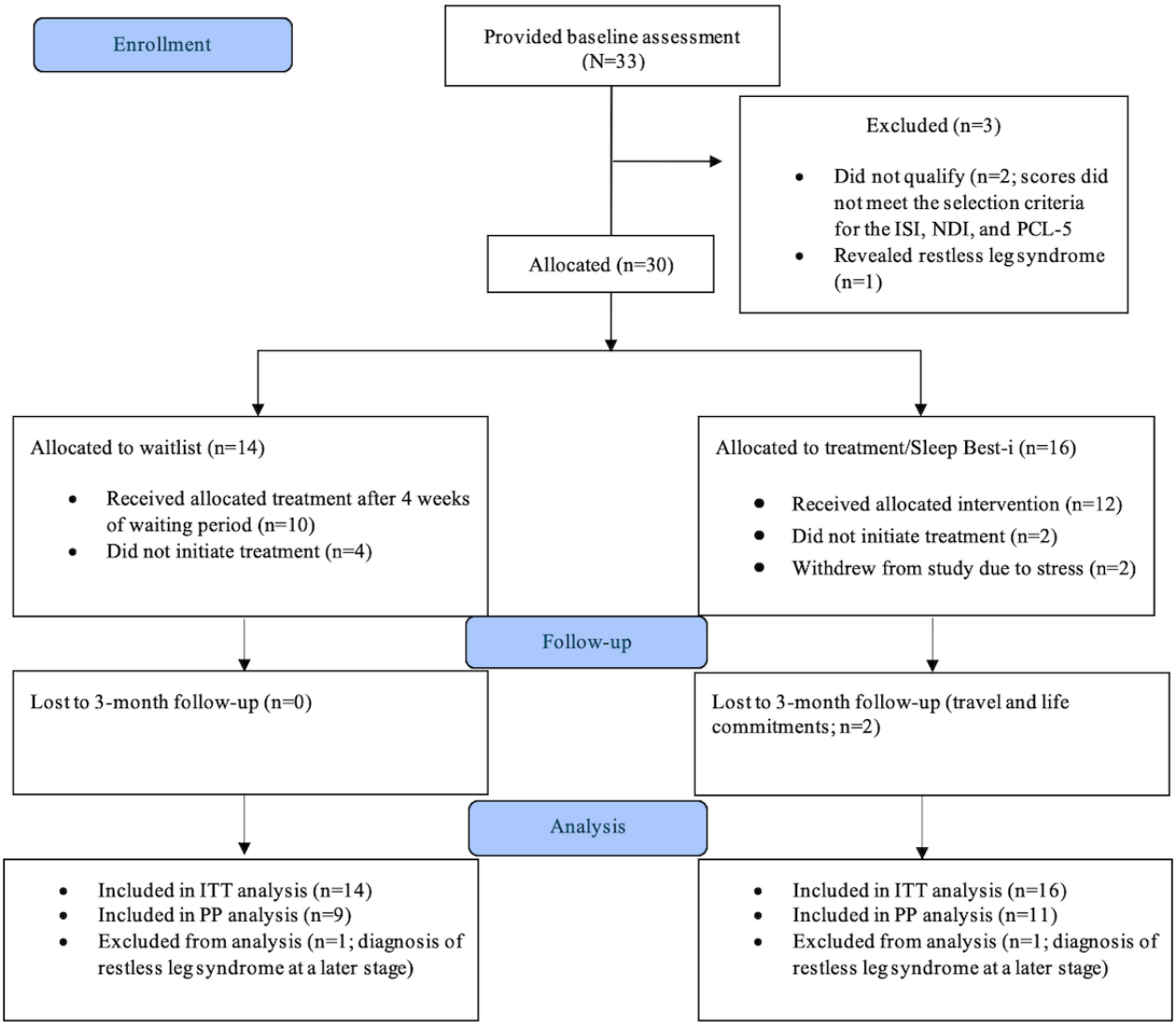


Table 2. Demographic variables for the treatment and waitlist groups and differences between the 2 at the baseline assessments.

Demographics	Treatment (n=16)	Waitlist (n=14)	Statistic	*P* value[a]
Age (y), mean (SD)	55.56 (15.14)	49.00 (17.34)	t_{28}=1.11	.28
Biological sex, n (%)			χ^2_2=1.4 (N=30)	.51
Female	11 (69)	8 (57)		
Male	5 (31)	6 (43)		
Country, n (%)			—[b]	—
Australia	12 (75)	14 (100)		
Canada	3 (19)	0 (0)		
United States	1 (6)	0 (0)		
Educational level, n (%)			χ^2_3=1.2 (N=30)	.75
Bachelor's degree	8 (50)	6 (43)		
Certificate or diploma	4 (25)	6 (43)		
High school	2 (12)	1 (7)		
Postgraduate	2 (12)	1 (7)		
Employment status, n (%)			χ^2_4=2.6 (N=30)	.64
Employed	7 (44)	5 (36)		
Looking for work	1 (6)	0 (0)		
Retired	5 (31)	5 (36)		
Student	2 (12)	1 (7)		
Unemployed	1 (6)	3 (21)		
Relationship status, n (%)			χ^2_3=8.0 (N=30)	.05
Married	8 (50)	5 (36)		
Divorced or separated	0 (0)	5 (36)		
Single	6 (38)	4 (29)		
Widowed	2 (12)	0 (0)		
Wildfires experienced, n (%)			—	—
2019-2020 wildfires (Australia)	8 (50)	11 (79)		
Waroona 2016 (Australia)	1 (6)	0 (0)		
Oregon 2020 (United States)	1 (6)	0 (0)		
2023 Canadian wildfires	1 (6)	0 (0)		
2021 BC fires	2 (12)	0 (0)		
Black Saturday 2009 (Australia)	2 (12)	1 (7)		
Currowan 2021 (Australia)	1 (6)	0 (0)		
Sacramento 2014 (United States)	0 (0)	1 (7)		
Blue Mountains 2013 (Australia)	0 (0)	1 (7)		

[a]Significance level at $P<.05$.

[b]Not applicable.

A total of 4 participants were excluded from the PP analysis for the following reasons: 2 (50%) participants, who completed the pilot trial and provided a complete set of data, revealed that they had restless leg syndrome that was not initially reported during the screening process; 1 (25%) participant received 1 treatment module but withdrew from the study due to illness, providing only baseline measures; and a fourth participant (25%) who provided and completed 50% of the data and modules withdrew due to illness and family commitments. Therefore, the PP analysis was conducted on a sample of 20 participants.

In total, 20% (6/30) of the participants provided baseline data; however, they did not initiate treatment and did not respond to emails sent by the research team. Table 1 indicates that there were no significant differences in demographic variables between the intervention and waitlist groups at the baseline assessment (age, *P*=.28; biological sex, *P*=.51; education level, *P*=.75; employment status, *P*=.64; and relationship status, *P*=.05).

To address hypothesis 1, ITT analysis of the mixed-effects regression, including the effect sizes of the DID in the fixed-effects component and the variability in the random-effects component, showed that, at baseline, the waitlist group had mean scores of 3.16, 0.39, and 45.29 on the ISI, NDI, and PCL-5, respectively. At the postintervention time point, the main effect of time, these scores significantly decreased by 1.64 points on the NDI (*P*=.001) and 10.64 points on the PCL-5 (*P*=.009), but no significant (*P*=.06) change was detected on the ISI. The treatment group showed slightly lower baseline scores than the waitlist group, with mean differences of 0.45, 0.04, and 9.20 for the ISI, NDI, and PCL-5, respectively, suggesting that the 2 groups were relatively balanced at baseline. The adjusted DID effects following the treatment indicated that the treatment group experienced greater improvements in mental health measures compared to the waitlist group, with differences of 1.54 points for the ISI, 1.22 points for the NDI, and 4.98 points for the PCL-5. Notably, the improvement in the NDI was marginally significant (*P*=.06), whereas the improvements in the ISI and PCL-5 were not statistically significant (*P*=.26 and *P*=.37, respectively; Table 3).

Table 3. Mixed-effects linear regression analysis of the effect of the intervention on primary outcomes (intention-to-treat analysis; N=30)[a].

Variable	Model 1 (n=60[b]; ISI[c])		Model 2 (n=60[b]; NDI[d])		Model 3 (n=60[b]; PCL-5[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	3.16 (−0.77 to 7.08)	.12	0.39 (−1.05 to 1.83)	.60	45.29 (37.27 to 53.30)	<.001
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−1.88 (−3.86 to 0.10)	.06	−1.64 (−2.58 to −0.71)	.001	−10.64 (−18.65 to −2.64)	.009
Condition						
Waitlist	Reference	—	Reference	—	Reference	—
Intervention	−0.45 (−2.56 to 1.66)	.68	−0.04 (−1.04 to 0.97)	.95	−9.20 (−20.17 to 1.78)	.10
Time × condition	−1.54 (−4.26 to 1.17)	.26	−1.22 (−2.50 to 0.06)	.06	−4.98 (−15.94 to 5.98)	.37
Random effects						
SD (time, intercept)	1.41 (0.82 to 2.42)	—	0.82 (0.55 to 1.20)	—	10.82 (7.26 to 16.14)	—
Correlation (time, intercept)	0.99 (−1.00 to 1.00)	—	1.00 (−1.00 to 1.00)	—	N/A[j]	—
SD of residual	2.48 (1.86 to 3.31)	—	1.23 (0.86 to 1.47)	—	10.81 (8.39 to 13.92)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue except for model 3, where the baseline measure was excluded for simplicity and convergence.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]ISI: Insomnia Severity Index.

[d]NDI: Nightmare Disorder Index.

[e]PCL-5: PTSD Checklist for DSM-5.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.03 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.001 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.19 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable as the random-intercept model better fit the data than the random-intercept, random-slope model. Given its better fit, random-intercept model parameters were reported.

In the random-effects part, there was substantial variability among individuals, with average deviations of 1.41, 0.82, and 10.82 points for the ISI, NDI, and PCL-5, respectively, around both the baseline and postintervention means. This suggests considerable individual variability in the outcomes. The positive correlation coefficients between the time point and the intercept indicate that participants with higher baseline scores (above the overall sample mean) were more likely to experience greater decreases in their scores over time compared to those with lower baseline scores. All mixed-effects models for the ISI, NDI, and PCL-5 were statistically significant (*P*<.001). Specifically, the random-intercept, random-slope model was appropriate for the

ISI (*P*=.03) and NDI (*P*=.001), whereas a random intercept–only model was sufficient for the PCL-5, which also fit the data but with a nonsignificant *P* value (*P*>.05 and *P*=.19; Table 3). Figure 2 illustrates the changes in primary outcome measures (ISI, NDI, and PCL-5) from before to after treatment based on the ITT analysis.

For completers only, PP analysis of the mixed-effects regression showed that, at baseline, the waitlist group had mean scores of 5.58, 0.46, and 13.37 on the ISI, NDI, and PCL-5, respectively. At the postintervention time point, these scores did not significantly change for the 3 measures (ISI; *P*=.68, NDI; *P*=.24, PCL-5; and *P*=.35). The treatment group baseline scores were balanced compared to those of the waitlist group, with mean differences of 0.99, 0.04, and 3.05 for the ISI, NDI, and PCL-5, respectively. However, the adjusted DID effects indicated an interaction effect of condition × time, with the treatment group experiencing greater and more significant improvements on the NDI with a decrease of 2.27 points (*P*=.049) and a decrease of 13.46 points on the PCL-5 (*P*=.03) but no significant improvements on the ISI (*P*=.25; Table 4).

There was a substantial variability among individuals, with average deviations of 3.41, 1.81, and 9.42 points for the ISI, NDI, and PCL-5, respectively, at the postintervention time point, suggesting considerable individual variability in the outcomes. All mixed-effects models for the ISI, NDI, and PCL-5 were statistically significant (*P*<.001), with a random-intercept model being appropriate for the 3 measures (Table 4).

In relation to the second hypothesis, the ITT analysis showed that, at baseline, the waitlist group had mean scores of 1.66, 12.01, and 0.69 on the GAD-7, PHQ-9, and PSQI, respectively (Table 5). Only the depression scores decreased significantly (*P*<.001) by 6.48 points at the posttreatment time point (main effect of time) in comparison to the waitlist group. The treatment group showed slightly lower baseline scores than those of the waitlist group on the secondary measures, with mean differences of 0.47, 0.90, and 0.04 on the GAD-7, PHQ-9, and PSQI, respectively, suggesting that the 2 groups were balanced at baseline. The adjusted effects at the postintervention time point indicate that the treatment group experienced greater improvements than the waitlist group on all measures but with only significant results on the PSQI (with a difference of 2.22 points; *P*<.001).

There was variability among individuals, as shown in Table 5, with average deviations of 0.91, 4.89, and 0.73 points for the GAD-7, PHQ-9, and PSQI, respectively, around both the baseline and postintervention means. The positive correlation coefficients for the GAD-7 and PSQI between the time point and the intercept indicate that participants with higher baseline scores (above the overall sample mean) were more likely to experience greater decrease in their scores over time compared to those with lower baseline scores. Mixed-effects models for the 3 measures were statistically significant (*P*<.001). Specifically, the random-intercept, random-slope model was appropriate for the PHQ-9 and PSQI (*P*=.007 and *P*=.001, respectively), whereas a random intercept–only model was sufficient for the GAD-7 (*P*=.08). Figure 3 shows the difference in decrease in scores between the intervention and waitlist groups for the GAD-7, PHQ-9, and PSQI from the ITT analysis.

For those who completed the study, PP analysis showed that the waitlist mean baseline scores were 11.78, 12.22, and 13.00 on the GAD-7, PHQ-9, and PSQI, respectively. These scores decreased significantly only for the PHQ-9 measure by 6.22 points at the posttreatment time point (*P*<.001). There was also a main effect of condition on the GAD-7 measure with a decrease of 4.14 points (*P*=.04) for the treatment group and an interaction effect of time × condition on the PSQI (*P*=.02), suggesting that the treatment group experienced significantly better sleep quality than the waitlist group following the treatment (Table 6).

Table 6 also shows a great variability among individuals on the random effects, with average deviations of 3.69, 3.16, and 2.82 points for the GAD-7, PHQ-9, and PSQI, respectively, around both the baseline and postintervention means. This suggests substantial individual variability in the outcome measures. Mixed-effects models for the 3 measures were statistically significant (*P*<.001), with the random-intercept, random-slope model being appropriate for the PHQ-9 (*P*=.004), whereas a random intercept–only model was sufficient for the GAD-7 and the PSQI (*P*=.28 and *P*>.99, respectively; Table 6).

To address the third hypothesis, separate mixed-effects linear model analyses were conducted for each group (intervention and waitlist) on all primary and secondary outcome measures. The analysis incorporated data from the following time points: baseline, posttreatment time point, and 3-month follow-up for the intervention group and pretreatment time point, posttreatment time point, and 3-month follow-up for the waitlist group. ITT analysis for the intervention group showed a significant reduction in insomnia, nightmares, and trauma symptoms as measured using the ISI, NDI, and PCL-5 both at the posttreatment time point (*P*<.001) with an effect of 3.42, 2.86, and 15.62 points, respectively, and at the 3-month follow-up (*P*<.001) with an effect of 8.64, 4.20, and 22.83 points, respectively (Table 7).

A great variability among individuals was detected on the random effects, with average deviations of 1.56, 0.32, and 11.09 points for the ISI, NDI, and PCL-5, respectively, around both the postintervention and 3-month follow-up means. Mixed-effects models for the 3 measures were statistically significant (*P*<.001), with the random-intercept, random-slope model being appropriate for the ISI and the PCL-5 (*P*=007 and *P*=.02, respectively), whereas the random-intercept model was appropriate for the NDI (*P*=.16; Table 7). Figure 4 shows the decrease in scores over time at the postintervention time point and 3-month follow-up for the intervention group.

For those who completed the study, the PP analysis showed comparable results to those reported in the ITT analysis, with a significant reduction in symptoms at both the postintervention time point (ISI; *P*=.04, NDI; *P*<.001, PCL-5; *P*<.001) and 3-month follow-up (*P*<.001 in all cases). The adjusted effects showed a continued reduction in symptoms from the postintervention time point to the 3-month follow-up, with a difference of 7.81, 4.86, and 23.71 points on the ISI, NDI, and PCL-5, respectively (Table 8).

The analysis also showed a substantial variability among individuals, with average deviations of 1.64, 1.03, and 1.81 points for the ISI, NDI, and PCL-5, respectively, around both the baseline and postintervention means. The positive correlation coefficients between the time point and the intercept on the NDI indicate that participants with higher baseline scores (above the overall sample mean) were more likely to experience greater decreases in their scores over time compared to those with lower baseline scores. All mixed-effects models for the ISI, NDI, and PCL-5 were statistically significant (P<.001), with a random-intercept model being sufficient for all 3 measures with a nonsignificant P value (ISI, P=.18; NDI, P=.07; and PCL-5, P=.42, Table 8).

On the secondary measures, the ITT analysis showed that the intervention group experienced significant improvements on the PHQ-9 and PSQI at both the posttreatment time point and the 3-month follow-up (P<.001). No significant reduction in symptoms was observed for the GAD-7 at the posttreatment time point (P=.21; Table 9). However, all participants in the intervention group experienced a significant reduction in symptoms at the 3-month follow-up (P<.001) on the 3 measures, with adjusted effects showing a decrease of 4.05, 7.50, and 5.12 points on the GAD-7, PHQ-9, and PSQI, respectively. Table 9 also shows variability among individuals on the random effects, with average deviations of 0.001, 5.46, and 0.82 points for the GAD-7, PHQ-9, and PSQI, respectively, around the postintervention and 3-month means. The 3 models were statistically significant (P<.001), with the random-intercept, random-slope model being appropriate for the PHQ-9 and PSQI (P=.03 and P=.002, respectively) and the random-intercept model being appropriate for the GAD-7 (P=.34). Figure 5 shows a change in secondary outcomes over time in the intervention group (ITT analysis).

Comparable results were also detected on the PP analysis, with individuals experiencing significant reductions in symptoms following the intervention and also at the 3-month assessments (P<.001 in all cases) except for the GAD-7, which was found to be not significant at the postintervention time point (P=.29). Participants experienced a greater reduction in symptoms at the 3-month assessment, with adjusted effects indicating a reduction of 3.88, 6.98, and 4.62 points on the GAD-7, PHQ-9, and PSQI, respectively (Table 10). The 3 models were statistically significant (P<.001), with the random-intercept model being appropriate for the 3 measures (GAD-7, P=.32; PHQ-9, P=.56, and PSQI, P=.05).

Similarly, ITT analysis results of the mixed-effects regression showed that, at baseline, participants in the waitlist group had mean scores of 10.79, 5.42, and 13.93 on the ISI, NDI, and PCL-5, respectively. The adjusted effects at the postintervention time point indicated that the symptoms significantly reduced by 6.36, 5.78, and 19.37 points, respectively (P<.001), and a significant reduction of 8.21, 7.74, and 27.96 points was observed at the 3-month follow-up for the ISI, NDI, and PCL-5, respectively (P<.001; Table 11).

In the random-effects part, there was substantial variability among individuals, with average deviations of 9.80, 6.18, and 1.73 points for the ISI, NDI, and PCL-5, respectively, around both the postintervention and 3-month means. This suggests considerable individual variability in the outcomes. All mixed-effects models for the ISI, NDI, and PCL-5 were statistically significant (P<.001), with a random-intercept model being appropriate for the 3 measures with a nonsignificant P value (P>.05; ISI, P=.45; NDI, P=.59, PCL-5, P=14 Table 11). Figure 6 shows the changes in primary outcomes over time for the waitlist group at the postintervention time point and 3-month follow-up.

The PP analysis showed similar results, with significant improvements at both the postintervention and 3-month follow-up assessments on the ISI, NDI, and PCL-5 (P<.001 in all cases; Table 12). The mixed-effects models for the ISI, NDI, and PCL-5 were all statistically significant (P<.001). The random-intercept model was used for the 3 measures (P>.05; ISI, P>.99; NDI, P>.99; and PCL-5, P=.34, Table 12).

For the secondary measures, ITT analysis showed that the waitlist group had baseline mean scores of 0.78, 0.36, and 3.31 on the GAD-7, PHQ-9, and PSQI, respectively. The adjusted effects at the postintervention time point indicated that these scores decreased significantly by 4.72 and 5.07 points for the GAD-7 and the PSQI, respectively (P<.001). However, no significant changes were observed on the PHQ-9 at the postintervention time point. Furthermore, adjusted effects showed a significant reduction of 6.51 points on the GAD-7 and 5.00 points on the PSQI (P<.001) against the baseline but no changes on the PHQ-9 at the 3-month follow-up (Table 13).

There was variability among individuals, with deviations of 1.47, 0.06, and 0.92 points for the GAD-7, PHQ-9, and PSQI, respectively, around the postintervention and 3-month follow-up means. The positive correlation coefficients between the time point and the intercept indicate that participants with higher baseline scores (above the overall sample mean) were more likely to experience greater decreases in their scores over time compared to those with lower baseline scores. All mixed-effects models for the GAD-7, PHQ-9, and PSQI were statistically significant (P<.001). Specifically, the random-intercept, random-slope model was appropriate for the GAD-7 and PSQI (P<.001 and P=.02, respectively), whereas a random intercept–only model was sufficient for the PHQ-9, which also fit the data but with a nonsignificant P value (P>.05; P=.17, Table 13). Figure 7 shows the changes in secondary outcomes at the postintervention time point and at the 3-month follow-up for the waitlist group.

For those who completed the study, PP analysis showed similar results for the waitlist group, with significant reduction at the postintervention and 3-month assessments only for the GAD-7 and the PSQI (P<.001); however, no significant changes were detected for the PHQ-9 at either the postintervention time point or the 3-month follow-up (P=.61 and P=.92, respectively; Table 14). For those who experienced significant reductions, there was great variability, with deviations of 1.98 and 1.00 points for the GAD-7 and PSQI, respectively, around the postintervention and 3-month follow-up means. All mixed-effects models for the GAD—7, PHQ—9, and PSQI were statistically significant (P<.001). Specifically, the random-intercept model was appropriate for the 3 measures

with a nonsignificant *P* values (*P*>.05; GAD-7, *P*=.05, PHQ-9, *P*=.47, PSQI, *P*=.24, Table 14).

Figure 2. Change in adjusted estimates of primary measures, including the Insomnia Severity Index (ISI); Nightmare Disorder Index (NDI); and PTSD Checklist for DSM-5 (PCL-5), from before to after the intervention based on the intention-to-treat analysis.

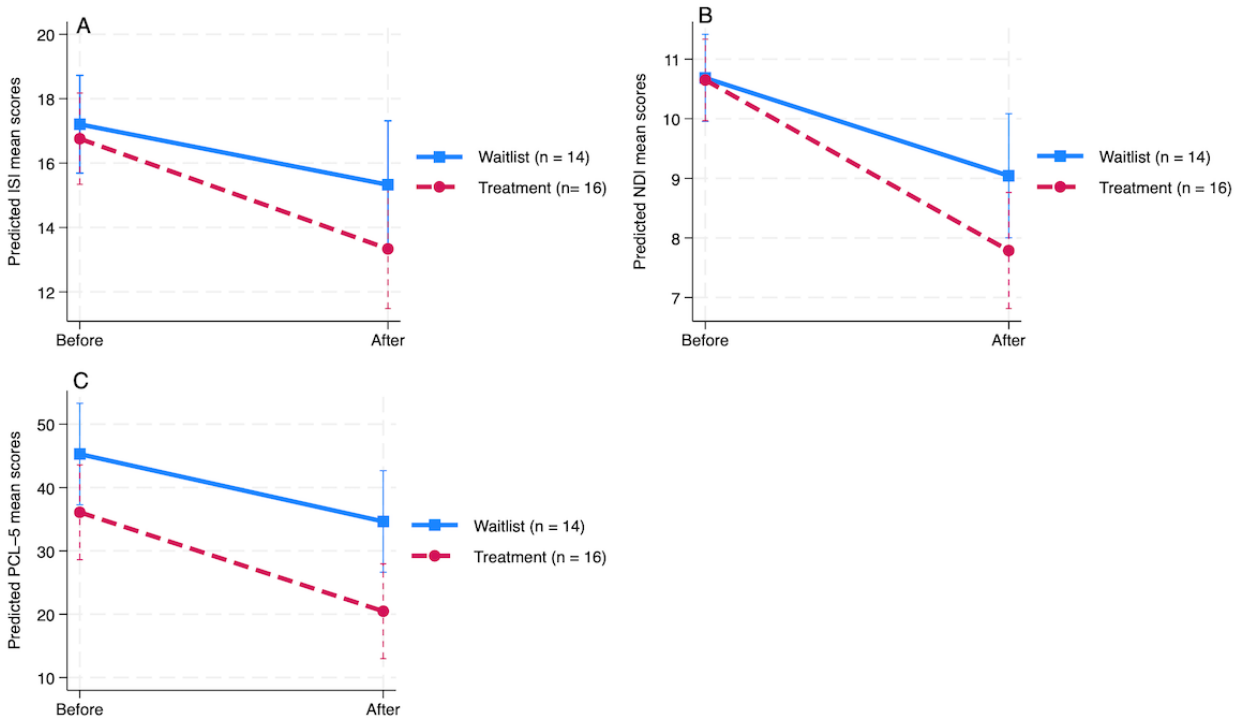


Table 4. Mixed-effects linear regression analysis of the effect of the intervention on primary outcomes for completers only (per-protocol analysis; N=20)[a].

Variable	Model 1 (n=40[b]; ISI[c])		Model 2 (n=40[b]; NDI[d])		Model 3 (n=40[b]; PCL-5[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	5.58 (1.12 to 10.04)	.01	0.46 (−1.70 to 2.61)	.68	13.37 (3.71 to 23.03)	.007
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−0.67 (−3.82 to 2.48)	.68	−1.00 (−2.67 to 0.67)	.24	−4.11 (−12.81 to 4.59)	.35
Condition						
Waitlist	Reference	—	Reference	—	Reference	—
Intervention	−0.99 (−4.06 to 2.04)	.53	−0.04 (−1.65 to 1.56)	.96	−3.05 (−11.52 to 5.42)	.48
Time × condition	−2.48 (−6.72 to 1.77)	.25	−2.27 (−4.53 to −0.01)	.049	−13.46 (−25.19 to −1.73)	.03
Random effects						
SD (time, intercept)	2.81 (2.60 to 3.22)	—	0.93 (1.31 to 2.00)	—	9.43 (6.67 to 10.33)	—
Correlation (time, intercept)	0.99 (-1.00 to 1.00)	—	1.00 (-1.00 to 1.00)	—	—	—
SD of residual	3.41 (2.74 to 4.24)	—	1.81 (1.46 to 2.26)	—	9.42 (7.57 to 11.73)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]ISI: Insomnia Severity Index.

[d]NDI: Nightmare Disorder Index.

[e]PCL-5: PTSD Checklist for DSM-5.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*>.99 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*>.99 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*>.99 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

Table 5. Mixed-effects linear regression analysis of the effect of the intervention on secondary outcomes (intention-to-treat analysis; N=30)[a].

Variable	Model 1 (n=60[b]; GAD-7[c])		Model 2 (n=60[b]; PHQ-9[d])		Model 3 (n=60[b]; PSQI[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	1.66 (−0.14 to 3.45)	.07	12.01 (9.46 to 14.57)	<.001	0.69 (−1.33 to 2.71)	.50
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	0.16 (−1.29 to 1.62)	.83	−6.48 (−9.04 to −3.92)	<.001	−0.58 (−1.43 to 0.27)	.18
Condition						
Waitlist	Reference	—	Reference	—	Reference	—
Intervention	−0.47 (−2.01 to 1.07)	.55	−0.90 (−4.40 to 2.61)	.62	−0.04 (−0.96 to 0.87)	.92
Time × condition	−1.62 (−3.62 to 0.38)	.11	1.28 (−2.23 to 4.78)	.48	−2.22 (−3.39 to −1.06)	<.001
Random effects						
SD (time and intercept)	0.91 (0.47 to 1.74)	—	4.89 (3.89 to 6.13)	—	0.73 (0.49 to 1.09)	—
Correlation (time and intercept)	N/A[j]	—	−0.79 (−0.89 to −0.61)	—	0.99 (−1.00 to 1.00)	—
SD of residual	1.86 (1.39 to 2.50)	—	0.000065 (0.000000665 to 0.01)	—	1.03 (0.78 to 1.34)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue except for model 2, where the baseline measure was excluded for simplicity and convergence.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PHQ-9: Patient Health Questionnaire–9.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.08 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.007 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.001 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable as the random-intercept model better fit the data than the random-intercept, random-slope model. Given its better fit, random-intercept model parameters were reported.

Figure 3. Change in adjusted estimates of secondary outcomes, including the Generalized Anxiety Disorder–7 (GAD-7), Patient Health Questionnaire–9 (PHQ-9), and Pittsburgh Sleep Quality Index (PSQI), from before to after the intervention based on the intention-to-treat analysis.

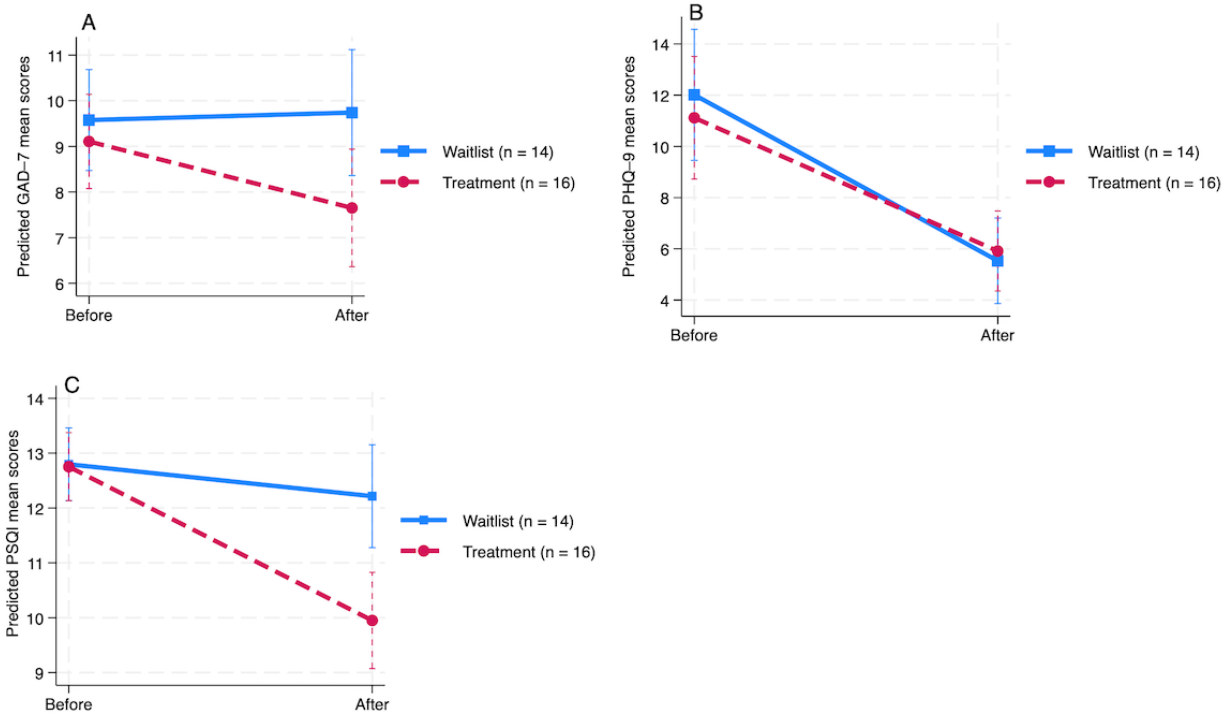


Table 6. Mixed-effects linear regression analysis of the effect of the intervention on secondary outcomes for completers only (per-protocol analysis; N=20)[a].

Variable	Model 1 (n=40[b]; GAD-7[c])		Model 2 (n=40[b]; PHQ-9[d])		Model 3 (n=40[b]; PSQI[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	11.78 (8.89 to 14.66)	<.001	12.22 (9.44 to 15.00)	<.001	13.00 (10.87 to 15.13)	<.001
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	1.33 (−0.90 to 3.57)	.24	−6.22 (−8.86 to −3.59)	<.001	−0.67 (−2.19 to 0.86)	.39
Condition						
Waitlist	Reference	—	Reference	—	Reference	—
Intervention	−4.14 (−8.03 to −0.25)	.04	−1.60 (−5.35 to 2.15)	.40	−1.00 (−3.88 to 1.88)	.50
Time × condition	−2.70 (−5.71 to 0.32)	.08	1.14 (−2.41 to 4.70)	.53	−2.42 (−4.48 to −0.37)	.02
Random effects						
SD (intercept)	3.69 (2.52 to 5.41)	—	3.16 (2.01 to 4.98)	—	2.82 (1.95 to 4.07)	—
Correlation (time, intercept)	N/A[j]	—	−0.68 (−0.85 to −0.37)	—	N/A	—
SD of residual	2.42 (1.78 to 3.30)	—	2.85 (2.09 to 3.89)	—	1.65 (1.21 to 2.25)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PHQ-9: Patient Health Questionnaire–9.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.28 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.004 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*>.99 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable as the random-intercept model better fit the data than the random-intercept, random-slope model. Given its better fit, random-intercept model parameters were reported.

Table 7. Mixed-effects linear regression analysis of the change in primary outcomes over time in the intervention group (intention-to-treat analysis; N=16)[a].

Variable	Model 1 (n=48[b]; ISI[c])		Model 2 (n=48[b]; NDI[d])		Model 3 (n=48[b]; PCL-5[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	3.23 (−1.66 to 8.12)	.20	3.37 (1.12 to 5.61)	.003	−17.52 (−27.73 to −7.30)	.001
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−3.42 (−5.47 to −1.38)	.001	−2.86 (−4.51 to −1.21)	.001	−15.62 (−21.87 to −9.37)	<.001
3 months after the intervention	−8.64 (−10.68 to −6.60)	<.001	−4.20 (−5.85 to −2.55)	<.001	−22.83 (−29.08 to −16.58)	<.001
Random effects						
SD (time, intercept)	1.56 (0.91 to 2.67)	—	0.32 (0.0000751 to 0.000133)	—	11.09 (7.06 to 17.43)	—
Correlation (time, intercept)	0.99 (−1.00 to 1.00)	—	N/A[j]	—	0.99 (−1.00 to 1.00)	—
SD of residual	2.73 (2.11 to 3.55)	—	2.38 (1.86 to 3.04)	—	4.45 (3.36 to 5.89)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]ISI: Insomnia Severity Index.

[d]NDI: Nightmare Disorder Index.

[e]PCL-5: PTSD Checklist for DSM-5.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.007 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.16 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.02 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable as the random-intercept model better fit the data than the random-intercept, random-slope model. Given its better fit, random-intercept model parameters were reported.

Figure 4. Change in Insomnia Severity Index (ISI); Nightmare Disorder Index (NDI); and PTSD Checklist for DSM-5 (PCL-5) scores for the intervention group from baseline to the postintervention time point and the 3-month follow-up as determined by the intention-to-treat analysis.

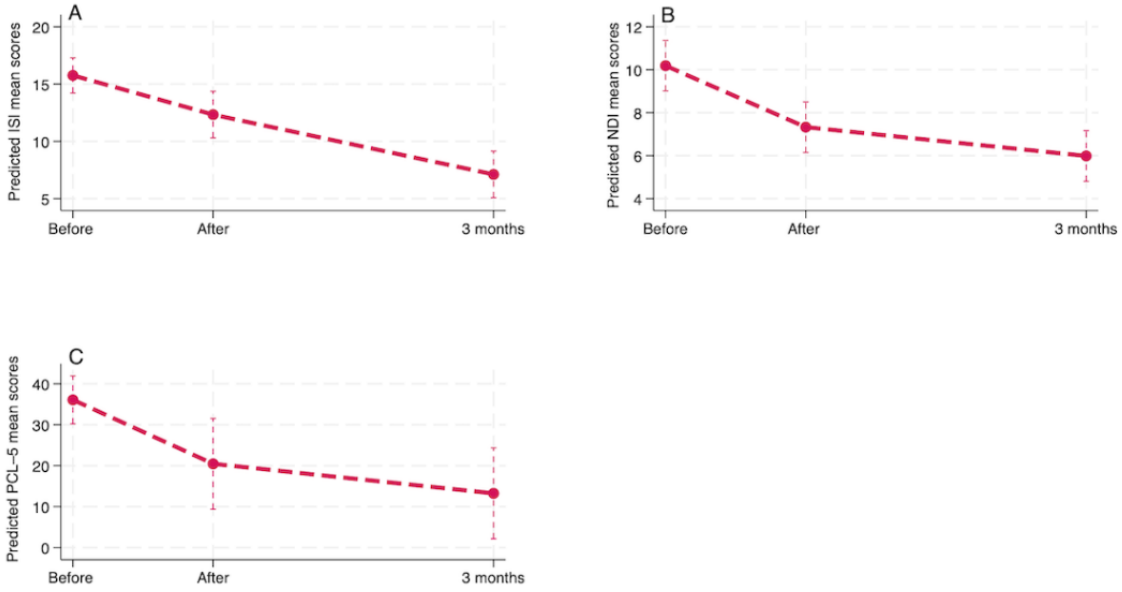


Table 8. Mixed-effects linear regression analysis of the change in primary outcomes over time in the intervention group for completers only (per-protocol analysis; N=11)[a].

Variable	Model 1 (n=33[b]; ISI[c])		Model 2 (n=33[b]; NDI[d])		Model 3 (n=33[b]; PCL-5[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	6.50 (−1.03 to 11.97)	.002	0.24 (−4.19 to 4.66)	.92	16.05 (7.49 to 24.61)	<.001
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−3.14 (−6.17 to −0.12)	.04	−3.27 (−4.66 to −1.89)	<.001	−17.57 (−25.45 to −9.69)	<.001
3 months after the intervention	−7.81 (−10.84 to −4.79)	<.001	−4.86 (−6.24 to −3.47)	<.001	−23.71 (−31.59 to −15.83)	<.001
Random effects						
SD (intercept)	1.64 (0.49 to 5.44)	—	1.03 (0.24 to 4.48)	—	1.81 (0.01 to 260.12)	—
SD of residual	3.62 (2.70 to 4.87)	—	1.49 (0.71 to 3.10)	—	9.43 (7.02 to 12.67)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]ISI: Insomnia Severity Index.

[d]NDI: Nightmare Disorder Index.

[e]PCL-5: PTSD Checklist for DSM-5.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.18 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.07 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.42 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

Table 9. Mixed-effects linear regression analysis of the change in secondary outcomes over time in the intervention group—intention-to-treat analysis (N=16)[a].

Variable	Model 1 (n=48[b]; GAD-7[c])		Model 2 (n=48[b]; PHQ-9[d])		Model 3 (n=48[b]; PSQI[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	3.64 (1.58 to 5.70)	.001	−8.64 (−12.69 to −4.59)	.001	0.92 (−2.45 to 4.29)	.59
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−1.46 (−3.73 to 0.82)	.21	−5.20 (−7.88 to −2.53)	<.001	−2.80 (−3.82 to −1.78)	<.001
3 months after the intervention	−4.05 (−6.32 to −1.77)	<.001	−7.50 (−10.17 to −4.82)	<.001	−5.12 (−6.14 to −4.11)	<.001
Random effects						
SD (time, intercept)	0.001 (0.0000198 to 0.0173)	—	5.46 (4.01 to 7.42)	—	0.82 (0.53 to 1.37)	—
Correlation (time, intercept)	N/A[j]	—	0.78 (0.54 to 0.90)	—	0.99 (−1.00 to 1.00)	—
SD of residual	3.28 (2.69 to 4.01)	—	0.001 (0.00000549 to 0.20)	—	1.34 (1.04 to 1.73)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PHQ-9: Patient Health Questionnaire–9.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.34 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.03 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.002 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable as the random-intercept model better fit the data than the random-intercept, random-slope model. Given its better fit, random-intercept model parameters were reported.

Figure 5. Change in Generalized Anxiety Disorder–7 (GAD-7), Patient Health Questionnaire–9 (PHQ-9), and Pittsburgh Sleep Quality Index (PSQI) scores for the intervention group from baseline to the postintervention time point and the 3-month follow-up as determined by the intention-to-treat analysis.

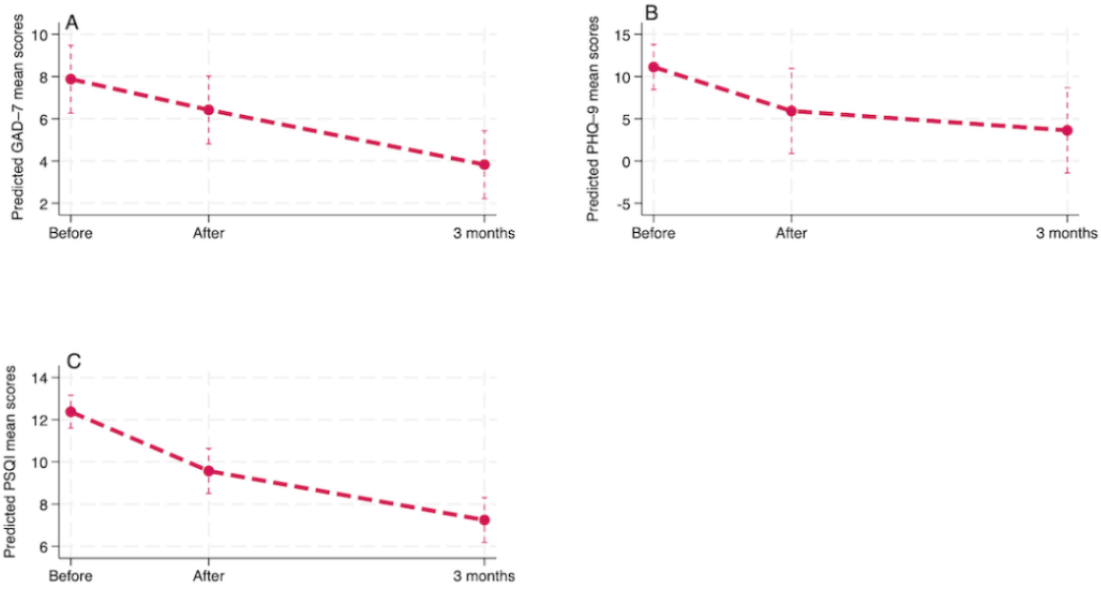


Table 10. Mixed-effects linear regression analysis of the change in secondary outcomes over time in the intervention group—per-protocol analysis (N=11)[a].

Variable	Model 1 (n=33[b]; GAD-7[c])		Model 2 (n=33[b]; PHQ-9[d])		Model 3 (n=33[b]; PSQI[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	2.85 (0.31 to 5.39)	.03	10.63 (8.28 to 12.98)	<.001	0.85 (−3.05 to 4.76)	.67
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−1.36 (−3.86 to 1.14)	.29	−5.08 (−7.80 to −2.36)	<.001	−3.09 (−4.50 to −1.68)	<.001
3 months after the intervention	−3.88 (−6.38 to −1.38)	.002	−6.98 (−9.71 to −4.26)	<.001	−4.62 (−6.03 to −3.21)	<.001
Random effects						
SD (intercept)	0.90 (0.10 to 8.32)	—	2.28 (1.100 to 4.73)	—	1.12 (0.52 to 2.41)	—
SD of residual	2.99 (2.23 to 4.02)	—	3.26 (2.42 to 4.38)	—	1.69 (1.26 to 2.27)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PHQ-9: Patient Health Questionnaire–9.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.32 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.56 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.05 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

Table 11. Mixed-effects linear regression analysis of the change in primary outcomes over time in the waitlist group receiving the intervention—intention-to-treat analysis (N=14)[a].

Variable	Model 1 (n=42[b]; ISI[c])		Model 2 (n=42[b]; NDI[d])		Model 3 (n=42[b]; PCL-5[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	10.79 (5.89 to 15.70)	<.001	5.42 (2.74 to 8.10)	<.001	13.93 (4.95 to 22.92)	.002
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−6.36 (−9.16 to −3.57)	<.001	−5.78 (−8.02 to −3.53)	<.001	−19.37 (−26.62 to −12.13)	<.001
3 months after the intervention	−8.21 (−11.01 to −5.41)	<.001	−7.74 (−9.98 to −5.49)	<.001	−27.96 (−35.20 to −20.71)	<.001
Random effects						
SD (intercept)	0.000098 (−0.21 to 0.21)	—	0.0000618 (−0.14 to 0.14)	—	1.73 (0.01 to 302.24)	—
SD of residual	3.78 (3.05 to 4.68)	—	3.03 (2.44 to 3.75)	—	9.78 (7.53 to 12.71)	—

[a]As the random-intercept model better fit the data than the random-intercept, random-slope model, random-intercept model parameters were reported. All models were adjusted for the pretreatment measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]ISI: Insomnia Severity Index.

[d]NDI: Nightmare Disorder Index.

[e]PCL-5: PTSD Checklist for DSM-5.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.45 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.59 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.14 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

Figure 6. Change in Insomnia Severity Index (ISI); Nightmare Disorder Index (NDI); and PTSD Checklist for DSM-5 (PCL-5) scores for the waitlist group from before to after the intervention and the 3-month follow-up as determined by the intention-to-treat analysis.

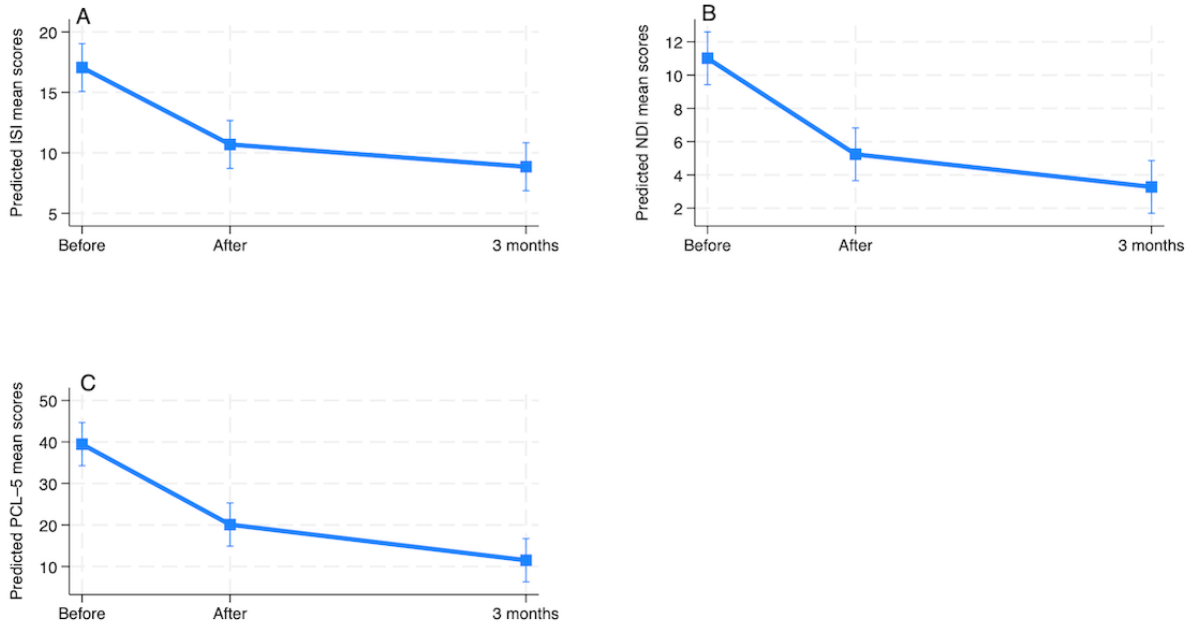


Table 12. Mixed-effects linear regression analysis of the change in primary outcomes over time in the waitlist group receiving the intervention—per-protocol analysis (N=9)[a].

Variable	Model 1 (n=27[b]; ISI[c])		Model 2 (n=27[b]; NDI[d])		Model 3 (n=27[b]; PCL-5[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	11.65 (5.59 to 17.71)	<.001	5.06 (1.47 to 8.66)	.006	12.85 (−1.41 to 27.11)	.08
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−6.56 (−10.49 to −2.63)	.001	−5.44 (−8.59 to −2.30)	.001	−20.89 (−30.98 to −10.79)	<.001
3 months after the intervention	−7.00 (−10.93 to −3.07)	<.001	−7.11 (−10.26 to −3.96)	<.001	−28.44 (−38.54 to −18.35)	<.001
Random effects						
SD (intercept)	0.01 (0.01 to 0.6)	—	0.01 (N/A[j])	—	3.31 (0.29 to 38.15)	—
SD of residual	4.25 (3.26 to 5.55)	—	3.41 (2.61 to 4.45)	—	10.93 (7.89 to 15.15)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]ISI: Insomnia Severity Index.

[d]NDI: Nightmare Disorder Index.

[e]PCL-5: PTSD Checklist for DSM-5.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*>.99 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*>.99 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.34 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable.

Table 13. Mixed-effects linear regression analysis of the change in secondary outcomes over time in the waitlist group receiving the intervention—intention-to-treat analysis (N=14)[a].

Variable	Model 1 (n=42[b]; GAD-7[c])		Model 2 (n=42[b]; PHQ-9[d])		Model 3 (n=42[b]; PSQI[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	0.78 (−2.68 to 4.24)	.66	0.36 (−1.00 to 1.72)	.61	3.31 (–0.62 to 7.25)	.10
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−4.72 (−6.24 to −3.19)	<.001	−0.42 (−1.84 to 1.00)	.56	−5.07 (−6.42 to −3.71)	<.001
3 months after the intervention	−6.51 (−8.03 to −4.99)	<.001	−0.40 (−1.82 to 1.02)	.58	−5.00 (−6.35 to −3.65)	<.001
Random effects						
SD (time, intercept)	1.47 (0.87 to 2.48)	—	0.06 (2.72×10^{-64} to 1.3×10^{61})	—	0.92 (0.48 to 1.77)	—
Correlation (time, intercept)	0.99 (−1.00 to 1.00)	—	N/A[j]	—	0.99 (−1.00 to 1.00)	—
SD of residual	1.78 (1.33 to 2.37)	—	1.91 (1.48 to 2.48)	—	1.71 (1.28 to 2.78)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PHQ-9: Patient Health Questionnaire–9.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*<.001 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.17 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.02 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

[j]N/A: not applicable as the random-intercept model better fit the data than the random-intercept, random-slope model. Given its better fit, random-intercept model parameters were reported.

Figure 7. Change in Generalized Anxiety Disorder–7 (GAD-7), Patient Health Questionnaire–9 (PHQ-9), and Pittsburgh Sleep Quality Index (PSQI) scores for the waitlist group from before to after the intervention and the 3-month follow-up as determined by the intention-to-treat analysis.

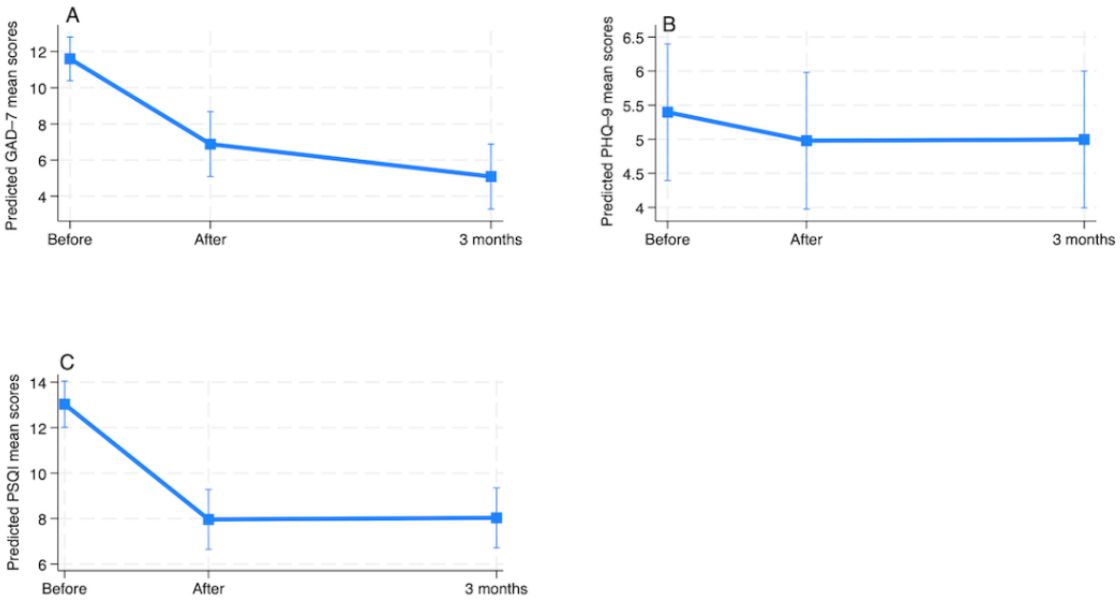


Table 14. Mixed-effects linear regression analysis of the change in secondary outcomes over time in the waitlist group receiving the intervention—per-protocol analysis (N=9)[a].

Variable	Model 1 (n=27[b]; GAD-7[c])		Model 2 (n=27[b]; PHQ-9[d])		Model 3 (n=27[b]; PSQI[e])	
	β (95% CI)	*P* value[f]	β (95% CI)	*P* value[g]	β (95% CI)	*P* value[h]
Fixed effects						
Intercept	3.85 (−8.29 to −3.04)	.22	0.43 (−1.61 to 2.48)	.68	5.89 (1.55 to 10.22)	.008
Time						
Baseline	Reference	—[i]	Reference	—	Reference	—
Postintervention time point	−5.67 (−8.29 to −3.04)	<.001	−0.56 (−2.69 to 1.58)	.61	−4.78 (−7.05 to −2.50)	<.001
3 months after the intervention	−7.78 (−10.40 to −5.15)	<.001	−0.11 (−2.25 to 2.03)	.92	−4.78 (−7.05 to −2.50)	<.001
Random effects						
SD (intercept)	1.98 (0.88 to 4.46)	—	0.29 (0.0000012 to 69,857.52)	—	1.00 (0.21 to 4.67)	—
SD of residual	2.84 (2.05 to 3.94)	—	2.31 (1.67 to 3.21)	—	2.46 (1.78 to 3.41)	—

[a]All models were adjusted for the baseline measure to address the regression-to-mean issue.

[b]*n* represents the number of observations in the data with long format to support mixed-effects modeling.

[c]GAD-7: Generalized Anxiety Disorder–7.

[d]PHQ-9: Patient Health Questionnaire–9.

[e]PSQI: Pittsburgh Sleep Quality Index.

[f]Model fit—*P*<.001 (Wald chi-square test) and *P*=.05 (likelihood ratio test [LRT]) to compare the mixed-effects models with random-intercept linear models.

[g]Model fit—*P*<.001 (Wald chi-square test) and *P*=.47 (LRT) to compare the mixed-effects models with random-intercept linear models.

[h]Model fit—*P*<.001 (Wald chi-square test) and *P*=.24 (LRT) to compare the mixed-effects models with random-intercept linear models.

[i]Not available.

Clinical Significance

Table 15 presents the number of participants achieving clinical significance from baseline to 3-month follow-up, based on intention-to-treat (ITT) analysis. The results show that a higher percentage of waitlist group participants achieved clinically significant changes in nightmare symptoms (mean change 7.74, SD 5.12) and PTSD symptoms (mean difference 27.96, SD 16.56) compared to the treatment group (mean change 4.20, SD 4.21 for nightmare symptoms and mean change: 22.83, SD 17.09 for PTSD symptoms). In contrast, a higher percentage of treatment group participants achieved clinically significant changes in insomnia symptoms (mean difference 8.65, SD 5.28) compared to the waitlist group (mean change 8.21, SD 6.96). Two participants were missing 3-month follow-up data on the ISI, NDI, and PCL-5.

The per-protocol (PP) analysis (Table 16) yielded similar results to the ITT analysis. Specifically, a higher percentage of waitlist group participants achieved clinically significant changes in nightmare symptoms (mean change 7.11, SD 5.71) and PTSD symptoms (mean change 28.44, SD 18.58) compared to the treatment group (mean change 4.67, SD 3.04 for nightmare symptoms and mean change: 24.84, SD 18.62 for PTSD symptoms). Conversely, a higher percentage of treatment group participants achieved clinically significant changes in insomnia symptoms (mean change 7.33, SD 5.87) compared to the waitlist group (mean change 7.00, SD 8.34). However, the differences between the two groups were not statistically significant.

Table 15. Distribution of clinically significant responders by condition—intention-to-treat analysis (N=30).

Outcome	Waitlist (n=14), n (%)	Treatment (n=16), n (%)	*P* value
ISI[a]	8 (57)	10 (62)	.77
NDI[b]	13 (93)	10 (62)	.05
PCL-5[c]	11 (79)	9 (56)	.20

[a]ISI: Insomnia Severity Index.

[b]NDI: Nightmare Disorder Index.

[c]PCL-5: PTSD Checklist for DSM-5.

Table 16. Distribution of clinically significant responders by condition—per-protocol analysis (N=18) (N=20).

Outcome	Waitlist (n=9), n (%)	Treatment (n=9), n (%)	*P* value
ISI[a]	4 (44)	5 (56)	.64
NDI[b]	8 (89)	5 (56)	.11
PCL-5[c]	7 (78)	4 (44)	.15

[a]ISI: Insomnia Severity Index.

[b]NDI: Nightmare Disorder Index.

[c]PCL-5: PTSD Checklist for DSM-5.

Satisfaction and Engagement With Treatment

Participants were asked to rate how likely they were to revisit the treatment modules on a 5-point, single-item Likert scale. A frequency analysis showed that most participants responded with *strongly agree* (9/20, 45%) and *moderately agree* (7/20, 35%). In total, 10% (2/20) of the participants responded with *moderately disagree*, and 10% (2/20) of the participants did not respond to this question. The number of log-ins to the site was observed as an indicator of the level of engagement by participants with the modules, and it was found that the 20 participants who were included in the PP analysis had an average of 8.5 log-ins during their engagement with Sleep Best-i. A total of 5% (2/20) logged in 12 times to the site; 5% (3/20) logged in 5 times; 5% (2/20) logged in 7 times; 10% (3/20) logged in 6 times; 15% (4/20) logged in 11 times; and 20% (6/20) logged in 8, 9, and 10 times. On average, participants spent 116 minutes on the site visiting modules and completing assessments.

Discussion

Principal Findings

The aim of this clinical pilot trial was to assess the feasibility of a brief (6 modules over 4 weeks), self-paced, digital intervention for the treatment of insomnia, nightmares, and PTSD symptoms in an international sample of survivors of wildfires. The first hypothesis was partially supported. The PP analysis revealed a significant interaction effect of condition × time on both the NDI and the PCL-5, indicating that Sleep Best-i effectively reduced symptoms of nightmares and PTSD from before to after the intervention in the treatment group compared to the waitlist group. However, no significant changes were observed in insomnia symptoms. The ITT analysis yielded similar findings, with a significant main effect of time showing a reduction in nightmare and PTSD symptoms at the posttreatment time point for the intervention group but no significant changes in insomnia symptoms. In examining the 2 groups separately, Sleep Best-i significantly reduced symptoms of insomnia, nightmares, and PTSD from baseline to the postintervention time point, and this improvement in symptoms was maintained at the 3-month assessment for the 2 groups across both the PP and ITT analyses. This study's findings diverge from those of previous research using CBTi to treat insomnia in survivors of wildfires [27,28] and veterans [25,26]. This discrepancy may be attributed to the brief duration of insomnia treatment in our study, which spanned only the first 2 weeks, unlike in other clinical trials that used CBTi in ≥6 sessions. Research suggests that an average of 6 to 8 sessions is typically required to significantly reduce insomnia symptoms [13,64]. The shorter session duration used in this study may have contributed to the differing outcomes. Nevertheless, the maintenance of improvements at the 3-month follow-up is a promising indicator of the intervention's long-term effectiveness.

These findings further substantiate the efficacy of ERRT in reducing the severity and frequency of nightmares, aligning with those of previous studies [36,65-67]. These results are also consistent with those of other clinical trials demonstrating that CBTi and ERRT can lead to significant reductions in PTSD symptoms at the posttreatment time point [36,68]. Notably, it is possible that symptoms of trauma and nightmares are more malleable and responsive to treatment, whereas insomnia symptoms may be more entrenched and less amenable to change with brief interventions, suggesting a potential explanation for the observed differences in treatment outcomes.

In relation to the second hypothesis, the PP analysis revealed that Sleep Best-i not only alleviated symptoms of nightmares and trauma but also significantly reduced anxiety symptoms (main effect of condition) in the treatment group compared to the waitlist group at the postintervention time point. In addition, the treatment group experienced a significant reduction in depressive symptoms at the posttreatment time point (main effect of time) and a significant interaction effect of time × condition on sleep quality, indicating more pronounced improvements in sleep quality at the postintervention time point compared to the waitlist group. The ITT analysis yielded similar findings, with significant reductions in depressive symptoms at the postintervention time point and a significant interaction effect of time × condition on sleep quality. However, the reduction in anxiety symptoms was no longer significant in the ITT analysis. This is notable as research suggests a strong relationship between insomnia and anxiety symptoms, with the expectation that successful insomnia treatment would lead to a reduction in anxiety symptoms [69].

When analyzing the 2 groups separately, both the PP and ITT analyses yielded similar results, with few exceptions. Notably, both groups showed significant reductions in anxiety symptoms at the 3-month follow-up, although this improvement was only sustained for the waitlist group at the postintervention assessment. In terms of sleep quality, both groups showed significant improvements at both the postintervention and 3-month follow-up assessments. However, only the intervention group exhibited significant reductions in depressive symptoms at both the postintervention and 3-month follow-up assessments.

This study's findings align with those of existing research demonstrating the effectiveness of CBTi-based treatments for comorbid conditions [13,70]. While part of this study's findings corroborates those of previous studies showing a significant reduction in symptoms of depression following CBT treatment for insomnia [27,28], the lack of improvement in symptoms of depression in the waitlist group is not well understood. One possible explanation is that the insignificant improvements in insomnia symptoms may have contributed to the absence of change in depression levels [69]. In addition, external life difficulties may have influenced symptoms of depression during the study period. Notably, the 2 groups had distinct intervention experiences, which may have impacted the outcomes. Specifically, the 3-month follow-up assessment for the waitlist group occurred 4 weeks after the treatment group's follow-up, potentially introducing passage-of-time effects that may have influenced the results. This study's findings are also in line with those of other studies that found improved sleep quality on the PSQI following the administration of CBTi [71].

In terms of clinical significance, a substantial proportion of participants achieved MCSC (according to the PP analysis) in insomnia (9/18, 50%), nightmares (13/18, 72%), and PTSD symptoms (11/18, 61%). Notably, the MCSC rates for insomnia and PTSD in this study were slightly higher than those reported in previous clinical trials using CBTi and imagery rehearsal therapy. For example, Ulmer et al [26] found that 55.4% and 50% of their sample (N=22) achieved clinical significance in insomnia and PTSD, respectively. In contrast, the MCSC rates in this study were lower than those reported by Belleville et al [27], who found that 64.7% and 70.6% of participants achieved MCSC for insomnia and PTSD at the posttreatment time point, respectively, and 64.7% and 58.8% achieved MCSC for insomnia and PTSD at the 3-month follow-up, respectively. The discrepancy in MCSC rates between our study and others may be attributed to the shorter treatment duration of Sleep Best-i (4 weeks) than those of the other trials (6-12 weeks) [26,27].

A significant proportion of participants expressed a high level of satisfaction with the intervention, with approximately 80% (18/20) of users reporting a positive experience. Furthermore, the level of engagement by participants with the modules was measured using the number of log-ins into the site. The 20 participants included in the PP analysis had an average of 8.5 log-in times during their engagement with Sleep Best-i. The log-in frequency was varied, with 5% (2/20) of the participants logging in 12 times; 5% (3/20) logging in 5 times; 5% (2/20) logging in 7 times; 10% (3/20) logging in 6 times; 15% (4/20) logging in 11 times; and 20% (6/20) logging in 8, 9, or 10 times. On average, participants spent 116 minutes on the site visiting modules and completing assessments.

Limitations

This clinical pilot trial has limitations that warrant consideration. The sample consisted of self-selected individuals, and the absence of clinical assessments to confirm diagnoses of insomnia, nightmares, and PTSD introduces a potential source of bias. The small sample size is another impediment, underscoring the need for further testing with a larger and more diverse population to establish the external validity of the intervention. One major concern in relation to Sleep Best-i is the PTSD module and its potential for triggering trauma symptoms. Although this risk was anticipated, participants were provided with emergency numbers. Moreover, throughout the initial 2-week treatment period, participants received training in cognitive and behavioral strategies to effectively manage stress. It is important to note that our study was conducted as an open-label trial. Future research should prioritize addressing the identified gaps and limitations, thereby enhancing the validity and generalizability of the findings in this field.

Future Directions

This feasibility trial provides valuable insights into the effectiveness of the Sleep Best-i intervention, a CBT-based program for survivors of wildfires. While our findings are promising, we recognize that future research could further enhance our understanding of its efficacy and effectiveness. One potential direction for future research could involve stratifying the assessment of outcomes based on the type of survivor by dividing the population into subgroups based on specific characteristics, such as demographic characteristics such as age, sex, and income; life experiences, including previous trauma and social support; experiences with wildfires, such as severity of exposure and proximity to the fire; losses experienced as a result of wildfires, including property damage and loss of loved ones; social support from family, friends, and community resources; and institutional support provided during the recovery phase, including access to mental health services and financial assistance. By doing so, researchers can identify patterns and trends within each subgroup, compare outcomes between subgroups, and develop targeted interventions that address the unique needs of each subgroup. In addition, ecological assessments over longer time frames could provide a more comprehensive understanding of the intervention's impact in real-world settings. We propose that future studies consider these approaches to build on our findings and inform the development of more targeted and effective interventions for survivors of wildfires.

Conclusions

Taken together, these findings indicate that Sleep Best-i incorporating CBTi and ERRT improved nightmares, PTSD, sleep quality, and symptoms of depression from baseline to the posttreatment time point. This positive impact was sustained at the 3-month follow-up for the 2 groups, with some variations on anxiety and depression. Participants in the intervention group, when assessed separately, experienced improvements on all measures from before to after treatment and at the 3-month follow-up, with the exception of anxiety symptoms at the posttreatment time point. The waitlist group experienced a significant reduction in symptoms on all measures from before treatment to after treatment and at the 3-month follow-up except for symptoms of depression. This clinical trial is the first in the field of sleep disturbances to use a concise, digital, self-paced intervention over 4 weeks among survivors of wildfires.

Acknowledgments

The research team would like to thank Dr Samia R Toukhsati for reviewing the ethics application and the treatment manual and for her contribution to the design of the trial; Dr Mirella DiBenedetto for her involvement in the early stages of this research project; and Dr Abby Bloom, Dr Kate Brady, Mr John Richardson, Mr Adam Dent, Ms Mona Naeimi, and Mr Nathan Maddock for their efforts in recruiting participants. The authors would also like to thank Mr Owen Cole for his assistance with the HealthZone platform. FI is supported by an Australian Government Research Training Program Fee Offset scholarship administered through Federation University. This study received funding from Natural Hazards Research Australia through an awarded scholarship to FI.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to ethical restrictions, but are available from the corresponding author on reasonable request.

Authors' Contributions

FI contributed to conceptualization; methodology; validation; formal analysis; data curation; writing—original draft preparation, review, and editing; visualization; project administration; and funding acquisition. BK contributed to conceptualization, methodology, validation, data curation, writing—review and editing, visualization, and supervision. HN contributed to methodology, validation, formal analysis, writing—review and editing, and visualization. SW provided supervision and writing—review and editing. GAK was responsible for the conceptualization, methodology, validation, formal analysis, data curation, writing—review and editing, visualization, and supervision. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

FI, BK, and GAK are the primary creators of Sleep Best-i. All other authors declare no other conflicts of interest.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1327 KB - humanfactors_v12i1e65228_app1.pdf]

Multimedia Appendix 2
Plain-language statement.
[DOCX File , 21 KB - humanfactors_v12i1e65228_app2.docx]

Multimedia Appendix 3
Consent form.
[PDF File (Adobe PDF File), 72 KB - humanfactors_v12i1e65228_app3.pdf]

References

1. Mani ZA, Khorram-Manesh A, Goniewicz K. Global health impacts of wildfire disasters from 2000 to 2023: a comprehensive analysis of mortality and injuries. Disaster Med Public Health Prep 2024 Oct 30;18:e230. [doi: 10.1017/dmp.2024.150] [Medline: 39473373]
2. Gallanter T, Bozeman WP. Firefighter illnesses and injuries at a major fire disaster. Prehosp Emerg Care 2002;6(1):22-26. [doi: 10.1080/10903120290938724] [Medline: 11789645]
3. Kulig JC, Townshend I, Edge DS, Reimer B. Impacts of wildfires: the aftermath at individual and community level. Aust J Emerg Manage 2013 Jul;28(3):23-28 [FREE Full text]
4. Shusterman D, Kaplan JZ, Canabarro C. Immediate health effects of an urban wildfire. West J Med 1993 Feb;158(2):133-138 [FREE Full text] [Medline: 8434462]
5. Johnston DW, Önder YK, Rahman MH, Ulubaşoğlu MA. Evaluating wildfire exposure: using wellbeing data to estimate and value the impacts of wildfire. J Econ Behav Organ 2021 Dec;192:782-798. [doi: 10.1016/j.jebo.2021.10.029]
6. Malhi NK, Marwaha R. Running wild: the impact of wildfires on mental health. Psychiatric Times. 2023 Oct 12. URL: https://www.psychiatrictimes.com/view/running-wild-the-impact-of-wildfires-on-mental-health [accessed 2024-01-13]
7. To P, Eboreime E, Agyapong VI. The impact of wildfires on mental health: a scoping review. Behav Sci (Basel) 2021 Sep 21;11(9):126 [FREE Full text] [doi: 10.3390/bs11090126] [Medline: 34562964]
8. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Association; 2013.

9. Schoenfeld FB, Deviva JC, Manber R. Treatment of sleep disturbances in posttraumatic stress disorder: a review. J Rehabil Res Dev 2012;49(5):729-752. [doi: 10.1682/jrrd.2011.09.0164] [Medline: 23015583]
10. Swift KM, Thomas CL, Balkin TJ, Lowery-Gionta EG, Matson LM. Acute sleep interventions as an avenue for treatment of trauma-associated disorders. J Clin Sleep Med 2022 Sep 01;18(9):2291-2312 [FREE Full text] [doi: 10.5664/jcsm.10074] [Medline: 35678060]
11. Koren D, Arnon I, Lavie P, Klein E. Sleep complaints as early predictors of posttraumatic stress disorder: a 1-year prospective study of injured survivors of motor vehicle accidents. Am J Psychiatry 2002 May;159(5):855-857. [doi: 10.1176/appi.ajp.159.5.855] [Medline: 11986142]
12. Isaac F, Toukhsati SR, Klein B, DiBenedetto M, Kennedy GA. Prevalence and predictors of sleep and trauma symptoms in wildfire survivors. Sleep Epidemiol 2023 Dec;3:100052. [doi: 10.1016/j.sleepe.2022.100052]
13. Isaac F, Toukhsati SR, DiBenedetto M, Kennedy GA. Cognitive behavioral therapy-based treatments for insomnia and nightmares in adults with trauma symptoms: a systematic review. Curr Psychol 2022 Jul 21:1-11 (forthcoming) [FREE Full text] [doi: 10.1007/s12144-022-03512-1] [Medline: 35891889]
14. Germain A, Richardson R, Moul DE, Mammen O, Haas G, Forman SD, et al. Placebo-controlled comparison of prazosin and cognitive-behavioral treatments for sleep disturbances in US Military Veterans. J Psychosom Res 2012 Feb;72(2):89-96 [FREE Full text] [doi: 10.1016/j.jpsychores.2011.11.010] [Medline: 22281448]
15. Taylor DJ, Peterson AL, Pruiksma KE, Young-McCaughan S, Nicholson K, Mintz J. Internet and in-person cognitive behavioral therapy for insomnia in military personnel: a randomized clinical trial. Sleep 2017 Jun 01;40(6). [doi: 10.1093/sleep/zsx075] [Medline: 28472528]
16. Walters EM, Jenkins MM, Nappi CM, Clark J, Lies J, Norman SB, et al. The impact of prolonged exposure on sleep and enhancing treatment outcomes with evidence-based sleep interventions: a pilot study. Psychol Trauma 2020 Feb;12(2):175-185 [FREE Full text] [doi: 10.1037/tra0000478] [Medline: 31246050]
17. Harb GC, Cook JM, Phelps AJ, Gehrman PR, Forbes D, Localio R, et al. Randomized controlled trial of imagery rehearsal for posttraumatic nightmares in combat veterans. J Clin Sleep Med 2019 May 15;15(5):757-767 [FREE Full text] [doi: 10.5664/jcsm.7770] [Medline: 31053215]
18. Ho FY, Chan CS, Tang KN. Cognitive-behavioral therapy for sleep disturbances in treating posttraumatic stress disorder symptoms: a meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:90-102. [doi: 10.1016/j.cpr.2015.09.005] [Medline: 26439674]
19. Morin CM, Espie CA. Insomnia: A Clinical Guide to Assessment and Treatment. Dordrecht, The Netherlands: Kluwer Academic Publishers; 2004.
20. Krakow B, Zadra A. Imagery rehearsal therapy: principles and practice. Sleep Med Clin 2010 Jun;5(2):289-298. [doi: 10.1016/j.jsmc.2010.01.004]
21. Davis JL. Treating Post-Trauma Nightmares: A Cognitive Behavioral Approach. Berlin, Germany: Springer; 2009.
22. Pruiksma KE, Wachen JS, Wardle S, Resick PA. Psychotherapy interventions for comorbid sleep disorders and posttraumatic stress disorder. In: Vermetten E, Germain A, Neylan T, editors. Sleep and Combat-Related Post Traumatic Stress Disorder. New York, NY: Springer; Jan 31, 2018.
23. Pruiksma KE, Taylor DJ, Mintz J, Nicholson KL, Rodgers M, Young-McCaughan S, et al. A pilot randomized controlled trial of cognitive behavioral treatment for trauma-related nightmares in active duty military personnel. J Clin Sleep Med 2020 Jan 15;16(1):29-40 [FREE Full text] [doi: 10.5664/jcsm.8116] [Medline: 31957648]
24. Taylor DJ, Peterson AL, Pruiksma KE, Hale WJ, Young-McCaughan S, Wilkerson A, et al. Impact of cognitive behavioral therapy for insomnia disorder on sleep and comorbid symptoms in military personnel: a randomized clinical trial. Sleep 2018 Jun 01;41(6). [doi: 10.1093/sleep/zsy069] [Medline: 29618098]
25. Margolies SO, Rybarczyk B, Vrana SR, Leszczyszyn DJ, Lynch J. Efficacy of a cognitive-behavioral treatment for insomnia and nightmares in Afghanistan and Iraq veterans with PTSD. J Clin Psychol 2013 Oct;69(10):1026-1042. [doi: 10.1002/jclp.21970] [Medline: 23629959]
26. Ulmer CS, Edinger JD, Calhoun PS. A multi-component cognitive-behavioral intervention for sleep disturbance in veterans with PTSD: a pilot study. J Clin Sleep Med 2011 Feb 15;07(01):57-68. [doi: 10.5664/jcsm.28042]
27. Belleville G, Ouellet MC, Békés V, Lebel J, Morin CM, Bouchard S, et al. Efficacy of a therapist-assisted self-help internet-based intervention targeting PTSD, depression, and insomnia symptoms after a disaster: a randomized controlled trial. Behav Ther 2023 Mar;54(2):230-246. [doi: 10.1016/j.beth.2022.08.004] [Medline: 36858756]
28. Krakow BJ, Melendrez DC, Johnston LG, Clark JO, Santana EM, Warner TD, et al. Sleep Dynamic Therapy for Cerro Grande Fire evacuees with posttraumatic stress symptoms: a preliminary report. J Clin Psychiatry 2002 Aug;63(8):673-684. [doi: 10.4088/jcp.v63n0804] [Medline: 12197447]
29. Rosenthal A, Stover E, Haar RJ. Health and social impacts of California wildfires and the deficiencies in current recovery resources: an exploratory qualitative study of systems-level issues. PLoS One 2021 Mar 26;16(3):e0248617 [FREE Full text] [doi: 10.1371/journal.pone.0248617] [Medline: 33770088]
30. Isaac F. Assessment of the effectiveness of online and face-to-face cognitive behavioural therapy for insomnia/nightmares in adults exposed to trauma using self-report and objective measures: preliminary findings. Trends Telemed E-Health 2022 Jan;3(2). [doi: 10.31031/TTEH.2022.03.000559]

31. Miller KE, Brownlow JA, Gehrman PR. Sleep in PTSD: treatment approaches and outcomes. Curr Opin Psychol 2020 Aug;34:12-17 [FREE Full text] [doi: 10.1016/j.copsyc.2019.08.017] [Medline: 31541965]
32. Mathe N, Johnson ST, Wozniak LA, Majumdar SR, Johnson JA. Alternation as a form of allocation for quality improvement studies in primary healthcare settings: the on-off study design. Trials 2015 Aug 25;16:375 [FREE Full text] [doi: 10.1186/s13063-015-0904-x] [Medline: 26303892]
33. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Trials 2010 Mar 24;11:32 [FREE Full text] [doi: 10.1186/1745-6215-11-32] [Medline: 20334632]
34. Lachenbruch PA, Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). J Am Stat Assoc 1989 Dec;84(408):1096. [doi: 10.2307/2290095]
35. Edinger JD. Treatment manual: cognitive-behavioral insomnia therapy. UNC School of Medicine. 2018. URL: https://www.med.unc.edu/neurology/wp-content/uploads/sites/716/2018/05/jdedingrCBTManual.pdf [accessed 2023-02-20]
36. Davis JL, Wright DC. Randomized clinical trial for treatment of chronic nightmares in trauma-exposed adults. J Trauma Stress 2007 Apr;20(2):123-133. [doi: 10.1002/jts.20199] [Medline: 17427914]
37. Edinger JD, Carney CE. Overcoming Insomnia: A Cognitive-Behavioral Therapy Approach, Therapist Guide. Oxford, UK: Oxford University Press; 2014.
38. Lynch J, Mack L, Benesek J, Marshall C, Clevinger L, McHenry S, et al. PTSD recovery program: treatment manual. Hunter Holmes McGuire VA Medical Center. 2015 Oct. URL: https://www.mirecc.va.gov/docs/visn6/PTSD_Recovery_Group-Client_Manual_3rd_edition.pdf [accessed 2025-02-25]
39. Germain A, Buysse DJ. Chapter 15 - Brief behavioral treatment of insomnia. In: Perlis M, Kuhn B, editors. Behavioral Treatments for Sleep Disorders: Practical Resources for the Mental Health Professional. Cambridge, MA: Academic Press; 2011:143-150.
40. Almatiana D, Astuti DS, Ramadhiyanti Y. A systematic review of sparkol videoscribe media in teaching writing skill. J Engl Lang Teach Educ 2021;2(1):1-15 [FREE Full text]
41. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
42. Morin CM, Belleville G, Bélanger L, Ivers H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. Sleep 2011 May 01;34(5):601-608 [FREE Full text] [doi: 10.1093/sleep/34.5.601] [Medline: 21532953]
43. Dietch JR, Taylor DJ, Pruiksma K, Wardle-Pinkston S, Slavish DC, Messman B, et al. The nightmare disorder index: development and initial validation in a sample of nurses. Sleep 2021 May 14;44(5):zsaa254 [FREE Full text] [doi: 10.1093/sleep/zsaa254] [Medline: 33245781]
44. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5) – standard [measurement instrument]. The National Centre for PTSD. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2025-02-25]
45. Krüger-Gottschalk A, Knaevelsrud C, Rau H, Dyer A, Schäfer I, Schellong J, et al. The German version of the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): psychometric properties and diagnostic utility. BMC Psychiatry 2017 Nov 28;17(1):379 [FREE Full text] [doi: 10.1186/s12888-017-1541-6] [Medline: 29183285]
46. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
47. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097 [FREE Full text] [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
48. Johnson SU, Ulvenes PG, Øktedalen T, Hoffart A. Psychometric properties of the General Anxiety Disorder 7-item (GAD-7) scale in a heterogeneous psychiatric sample. Front Psychol 2019 Aug 6;10:1713 [FREE Full text] [doi: 10.3389/fpsyg.2019.01713] [Medline: 31447721]
49. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
50. Buysse DJ, Reynolds CF3, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]
51. Backhaus J, Junghanns K, Broocks A, Riemann D, Hohagen F. Test-retest reliability and validity of the Pittsburgh Sleep Quality Index in primary insomnia. J Psychosom Res 2002 Sep;53(3):737-740. [doi: 10.1016/s0022-3999(02)00330-6] [Medline: 12217446]
52. StataCorp. Stata statistical software: release 18. StataCorp LLC. 2023. URL: https://www.stata.com/new-in-stata/ [accessed 2025-02-25]
53. Pallant J. SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS. London, UK: Routledge; Jul 16, 2020.

54. Hoaglin DC, Iglewicz B. Fine-tuning some resistant rules for outlier labeling. J Am Stat Assoc 1987;82(400):1147-1149. [doi: 10.1080/01621459.1987.10478551]
55. McCoy CE. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
56. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: intention-to-treat versus per-protocol analysis. Perspect Clin Res 2016;7(3):144-146 [FREE Full text] [doi: 10.4103/2229-3485.184823] [Medline: 27453832]
57. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? Int J Methods Psychiatr Res 2011 Mar;20(1):40-49 [FREE Full text] [doi: 10.1002/mpr.329] [Medline: 21499542]
58. von Hippel PT. How many imputations do you need? A two-stage calculation using a quadratic rule. Sociol Methods Res 2020 Aug;49(3):699-718 [FREE Full text] [doi: 10.1177/0049124117747303] [Medline: 39211325]
59. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. BMC Med Res Methodol 2014 Jun 05;14:75 [FREE Full text] [doi: 10.1186/1471-2288-14-75] [Medline: 24903709]
60. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:b2393 [FREE Full text] [doi: 10.1136/bmj.b2393] [Medline: 19564179]
61. Wing C, Yozwiak M, Hollingsworth A, Freedman S, Simon K. Designing difference-in-difference studies with staggered treatment adoption: key concepts and practical guidelines. Annu Rev Public Health 2024 May;45(1):485-505 [FREE Full text] [doi: 10.1146/annurev-publhealth-061022-050825] [Medline: 38277791]
62. Yang M, Morin CM, Schaefer K, Wallenstein GV. Interpreting score differences in the Insomnia Severity Index: using health-related outcomes to define the minimally important difference. Curr Med Res Opin 2009 Oct;25(10):2487-2494. [doi: 10.1185/03007990903167415] [Medline: 19689221]
63. Marx BP, Lee DJ, Norman SB, Bovin MJ, Sloan DM, Weathers FW, et al. Reliable and clinically significant change in the clinician-administered PTSD Scale for DSM-5 and PTSD Checklist for DSM-5 among male veterans. Psychol Assess 2022 Feb;34(2):197-203 [FREE Full text] [doi: 10.1037/pas0001098] [Medline: 34941354]
64. Pigeon WR. Treatment of adult insomnia with cognitive-behavioral therapy. J Clin Psychol 2010 Nov;66(11):1148-1160 [FREE Full text] [doi: 10.1002/jclp.20737] [Medline: 20853442]
65. Rhudy JL, Davis JL, Williams AE, McCabe KM, Byrd PM. Physiological-emotional reactivity to nightmare-related imagery in trauma-exposed persons with chronic nightmares. Behav Sleep Med 2008;6(3):158-177. [doi: 10.1080/15402000802162539] [Medline: 18629687]
66. Pruiksma KE, Cranston CC, Rhudy JL, Micol RL, Davis JL. Randomized controlled trial to dismantle exposure, relaxation, and rescripting therapy (ERRT) for trauma-related nightmares. Psychol Trauma 2018 Jan;10(1):67-75. [doi: 10.1037/tra0000238] [Medline: 27977223]
67. Balliett NE, Davis JL, Miller KE. Efficacy of a brief treatment for nightmares and sleep disturbances for veterans. Psychol Trauma 2015 Nov;7(6):507-515. [doi: 10.1037/tra0000055] [Medline: 26010113]
68. Talbot LS, Maguen S, Metzler TJ, Schmitz M, McCaslin SE, Richards A, et al. Cognitive behavioral therapy for insomnia in posttraumatic stress disorder: a randomized controlled trial. Sleep 2014 Feb 01;37(2):327-341 [FREE Full text] [doi: 10.5665/sleep.3408] [Medline: 24497661]
69. Mirchandaney R, Barete R, Asarnow LD. Moderators of cognitive behavioral treatment for insomnia on depression and anxiety outcomes. Curr Psychiatry Rep 2022 Feb;24(2):121-128 [FREE Full text] [doi: 10.1007/s11920-022-01326-3] [Medline: 35061137]
70. Sweetman A, Lovato N, Micic G, Scott H, Bickley K, Haycock J, et al. Do symptoms of depression, anxiety or stress impair the effectiveness of cognitive behavioural therapy for insomnia? A chart-review of 455 patients with chronic insomnia. Sleep Med 2020 Nov;75:401-410. [doi: 10.1016/j.sleep.2020.08.023] [Medline: 32957076]
71. Cook JM, Harb GC, Gehrman PR, Cary MS, Gamble GM, Forbes D, et al. Imagery rehearsal for posttraumatic nightmares: a randomized controlled trial. J Trauma Stress 2010 Oct;23(5):553-563. [doi: 10.1002/jts.20569] [Medline: 20839311]

Abbreviations

CBT: cognitive behavioral therapy
CBTi: cognitive behavioral therapy for insomnia
CONSORT: Consolidated Standards of Reporting Trials
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DID: difference-in-differences
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
ERRT: exposure, relaxation, and rescripting therapy
GAD-7: Generalized Anxiety Disorder–7
ISI: Insomnia Severity Index
ITT: intention-to-treat

MCSC: minimal clinically significant change
NDI: Nightmare Disorder Index
PCL-5: PTSD Checklist for DSM-5
PHQ-9: Patient Health Questionnaire–9
PLIS: plain-language information statement
PP: per protocol
PSQI: Pittsburgh Sleep Quality Index
PTSD: posttraumatic stress disorder

Edited by A Kushniruk; submitted 04.09.24; peer-reviewed by C Baxter; comments to author 15.01.25; revised version received 22.01.25; accepted 23.01.25; published 14.03.25.

Please cite as:
Isaac F, Klein B, Nguyen H, Watson S, Kennedy GA
Digital Cognitive Behavioral Therapy–Based Treatment for Insomnia, Nightmares, and Posttraumatic Stress Disorder Symptoms in Survivors of Wildfires: Pilot Randomized Feasibility Trial
JMIR Hum Factors 2025;12:e65228
URL: https://humanfactors.jmir.org/2025/1/e65228
doi:10.2196/65228
PMID:

Enhancing Methodological Rigor in Mobile Health Care Research

Shuhan Tang[1,2], MS

[1]Duke Global Health Institute, Duke University, Durham, NC, United States

[2]Global Health Research Center (GHRC), Duke Kunshan University, No. 8 Duke Avenue, Kunshan, China

Corresponding Author:
Shuhan Tang, MS
Duke Global Health Institute, Duke University, Durham, NC, United States

Related Article:

https://humanfactors.jmir.org/2024/1/e59153

(JMIR Hum Factors 2025;12:e68243) doi:10.2196/68243

KEYWORDS

letter to the editor; health care professionals; mobile health care; technical training; cross-sectional survey; mobile; China; web-based questionnaire; logistic regression; mhealth; mobile health

Tang et al's recent study published in *JMIR Human Factors* [1], titled "The Use of Mobile Health Care Among Medical Professionals in the Sichuan-Chongqing Region: Cross-Sectional Survey Study," captured my attention. Their analysis of mobile health (mHealth) device use and influencing factors, using chi-square and multivariable logistic regression analyses, revealed a significant association between age and mHealth use. This study provides valuable insights from China's western region.

However, I would like to offer a few comments and suggestions that I believe would further enhance the study's methodology and findings.

First, the study's reliance on a web-based questionnaire may introduce potential selection bias, as respondents, by nature of accessing web-based surveys, may be more likely to use digital devices and have a stronger interest in digital health in general. A more diverse, mixed-methods approach to questionnaire distribution could help mitigate this potential self-selection toward mHealth use and prevent the exclusion of individuals with limited experience or interest in mHealth.

Second, the study could benefit from a standardized and rigorous methodological framework for survey design and reporting. For instance, categorizing age into three broad groups may overlook essential trends. A finer measurement scale for satisfaction and usage, such as a Likert scale, could provide deeper insights into health care professionals' attitudes. Furthermore, for questions where "uncertain" responses outweigh "yes" or "no" responses, a qualitative or mixed-methods research approach could yield a more nuanced understanding of the underlying reasons. Finally, incorporating years of work experience as a variable could add valuable insights, given its potential correlation with age. Unlike professional titles, years of work experience could provide a more direct measure of professional tenure, potentially enriching the data analysis.

Third, ethical considerations of this study merit further attention. Identifying specific hospitals in the report may compromise confidentiality. Additionally, the exemption of review by an institutional review board (IRB) raises concerns, as the study involves gathering potentially sensitive information from human participants (ie, health care professionals) regarding personal perspectives, workflows, and technology use. Such studies often warrant an ethics board review to protect participant privacy, minimize psychological or social risks, and ensure adherence to ethical standards [2].

Fourth, before conducting logistic regression analysis, normality tests on both independent and dependent variables are recommended to confirm the validity of the chosen statistical methods [3]. If data are not normally distributed, adjustments should be considered.

Lastly, as noted in the "Limitations" section [1], the reliance on convenience sampling may affect the study's generalizability. Only public, district-level hospitals were included, excluding primary and tertiary institutions—this limits the representation of China's complete health care hierarchy in the analysis. Furthermore, only urban areas were sampled, overlooking rural populations. Women comprised 77.1% of the study sample, raising questions about gender representation. A multi-level sampling approach would likely yield a more comprehensive and representative dataset.

To conclude, I would like to highlight that these feedback points are not to challenge the integrity of the authors' work. Instead, I hope they can contribute to ongoing discussions on mHealth research and the development of robust methodologies in this field.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of the publication "The Use of Mobile Health Care Among Medical Professionals in the Sichuan-Chongqing Region: Cross-Sectional Survey Study" declined to respond to this letter.

References

1. Tang Y, Yang J, Wang N, Wang X, Hu W. The use of mobile health care among medical professionals in the Sichuan-Chongqing region: cross-sectional survey study. JMIR Hum Factors 2024 Oct 24;11:e59153. [doi: 10.2196/59153] [Medline: 39447153]
2. Beauchamp TL, Childress JF. Principles of Biomedical Ethics, 8th edition: Oxford University Press; 2019.
3. Field A. Discovering Statistics Using IBM SPSS Statistics, 4th edition: Sage Publications; 2013.

Abbreviations

IRB: institutional review board
mHealth: mobile health

Edited by A Iannaccio, T Leung; submitted 31.10.24; this is a non–peer-reviewed article;accepted 29.11.24; published 03.01.25.

Please cite as:
Tang S
Enhancing Methodological Rigor in Mobile Health Care Research
JMIR Hum Factors 2025;12:e68243
URL: https://humanfactors.jmir.org/2025/1/e68243
doi:10.2196/68243

Original Paper

Evaluating the Acceptance and Usability of an Independent, Noncommercial Search Engine for Medical Information: Cross-Sectional Questionnaire Study and User Behavior Tracking Analysis

Lisa Specht[1*], MD; Raphael Scheible[2*], MSc; Martin Boeker[2], Prof Dr Med; Erik Farin-Glattacker[3], Prof Dr; Nikolas Kampel[4], MSc; Marina Schmölz[1], MSc; Andrea Schöpf-Lazzarino[1,5], PhD; Stefan Schulz[6], Prof Dr Med; Christian Schlett[3], PhD; Fabian Thomczyk[7]; Sebastian Voigt-Radloff[3], PhD; Constanze Wegner[8], MA; Katharina Wollmann[8], MPH; Andy Maun[1], Prof Dr

[1]Institute of General Practice, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

[2]Institute of Artificial Intelligence and Informatics in Medicine, Chair of Medical Informatics, University Hospital rechts der Isar, School of Medicine and Health, Technical University of Munich, Munich, Germany

[3]Institute of Medical Biometry and Statistics, Section of Health Care Research and Rehabilitation Research, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany

[4]Institute of Neuroscience and Medicine, Jülich Aachen Research Alliance, Forschungszentrum Jülich GmbH, Jülich, Germany

[5]Careum School of Health, part of the Kalaidos University of Applied Sciences, Zurich, Switzerland

[6]Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria

[7]Data Integration Center, University of Freiburg, Freiburg, Germany

[8]Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany

[*]these authors contributed equally

Corresponding Author:
Lisa Specht, MD
Institute of General Practice
Faculty of Medicine and Medical Center
University of Freiburg
Elsässer Straße 2m
Freiburg, 79110
Germany
Phone: 49 761 270 72490
Email: lisa.specht@uniklinik-freiburg.de

Abstract

Background: The internet is a key source of health information, but the quality of content from popular search engines varies, posing challenges for users—especially those with low health or digital health literacy. To address this, the “tala-med” search engine was developed in 2020 to provide access to high-quality, evidence-based content. It prioritizes German health websites based on trustworthiness, recency, user-friendliness, and comprehensibility, offering category-based filters while ensuring privacy by avoiding data collection and advertisements.

Objective: This study aims to evaluate the acceptance and usability of this independent, noncommercial search engine from the users’ perspectives and their actual use of the search engine.

Methods: For the questionnaire study, a cross-sectional study design was used. In total, 802 participants were recruited through a web-based panel and were asked to interact with the new search engine before completing a web-based questionnaire. Descriptive statistics and multiple regression analyses were used to assess participants’ acceptance and usability ratings, as well as predictors of acceptance. Furthermore, from October 2020 to June 2021, we used the open-source web analytics platform Matomo to collect behavior-tracking data from consenting users of the search engine.

Results: The study indicated positive findings on the acceptance and usability of the search engine, with more than half of the participants willing to reuse (465/802, 58%) and recommend it (507/802, 63.2%). Of the 802 users, 747 (93.1%) valued the

absence of advertising. Furthermore, 92.3% (518/561), 93.9% (553/589), 94.7% (567/599), and 96.5% (600/622) of those users who used the filters agreed at least partially that the filter functions were helpful in finding trustworthy, recent, user-friendly, or comprehensible results. Participants criticized some of the search results regarding the selection of domains and shared ideas for potential improvements (eg, for a clearer design). Regression analyses showed that the search engine was especially well accepted among older users, frequent internet users, and those with lower educational levels, indicating an effective targeting of segments of the population with lower health literacy and digital health literacy. Tracking data analysis revealed 1631 sessions, comprising 3090 searches across 1984 unique terms. Users performed 1.64 (SD 1.31) searches per visit on average. They prioritized the search terms “corona,” “back pain,” and “cough.” Filter changes were common, especially for *recency* and *trustworthiness*, reflecting the importance that users placed on these criteria.

Conclusions: User questionnaires and behavior tracking showed the platform was well received, particularly by older and less educated users, especially for its advertisement-free design and filtering system. While feedback highlighted areas for improvement in design and filter functionality, the search engine’s focus on transparency, evidence-based content, and user privacy shows promise in addressing health literacy and navigational needs. Future updates and research will further refine its effectiveness and impact on promoting access to quality health information.

(JMIR Hum Factors 2025;12:e56941) doi:10.2196/56941

KEYWORDS

medical information; health information; search engine; user behavior; health literacy; digital health literacy; navigational needs; information-seeking behavior; evidence-based content; Germany

Introduction

Background

The internet has become a crucial resource for accessing medical information, making it a valuable tool for individuals seeking knowledge about health. In Germany, medical professionals, notably physicians, have long been the primary source of health care information for the general population [1,2]. However, the landscape is evolving, and the internet now stands as the second most significant channel for health-related information [1], indicating a new trend in how people seek medical knowledge. A study conducted by the Bertelsmann Foundation showed a strong and increasing demand for health-related information online, including a rapidly increasing demand by older generations [3]. In 2020, approximately 70% of the German population used the internet to actively engage in searching for health-related content [4]. While these numbers have varied in recent years, there is an overall trend of an increased use of the internet [4].

A major complaint that users have about health-related information online is the lack of clarity regarding the trustworthiness and seriousness of existing websites. The information online has varying levels of quality [5,6]. Frequently used and popular websites are regarded as reliable by users, irrespective of the actual quality of their content, while independent public websites are not as well known and do not seem to be more reliable to users [3]. The reliance on digital platforms for health information also poses a challenge due to the profound influence of the order of the results generated by popular search engines. Users often initiate their search for medical information through well-known search engines, predominantly relying on the first few search results displayed on the search engine results page (SERP), which are often advertising [7-10]. Unfortunately, these top hits are not only clicked on more frequently but are also perceived as more trustworthy, despite the variation in information quality [3,9]. This highlights the ambiguity of users feeling incompetent in finding reliable information online, while at the same time trusting the top results and being satisfied with the information they find. A more recent study evaluated the quality of online health and nutrition information related to cancer supplements through a Google search, using the Health Information Quality Index [11]. The 160 relevant search results yielded median and mean Health Information Quality Index scores of 8, with one-quarter of the results scoring high (10-12). No correlation was found between high quality scores and an early appearance of these results, indicating potential limitations in using Google for obtaining accurate information on dietary supplements and cancer, particularly given the prevalence of advertisements outnumbering search results [11].

The problem with the reliance on search engines becomes even clearer when considering the low health literacy and digital health literacy of the German population [12-14]. The first Health Literacy Survey Germany study in 2014 [13] underscored a significant health literacy challenge in Germany, with 54.3% of the population reporting health literacy problems. This result pointed to challenges in navigating the health care system and understanding health-related information, with identified associations with factors such as age, migrant background, self-assessed social status, and functional literacy. The second Health Literacy Survey Germany study in 2020 [14] reinforced previous concerns with a further increased rate of individuals (58.8%) struggling considerably with health information. Even more so than in 2014, evaluating the trustworthiness of health-related information presented difficulties [15]. Up to 45% of the German population exhibit “problematic” or even “inadequate” internet skills [16]. Many individuals face so-called navigational needs. They depend on support from others to efficiently search and evaluate health information online [17]. As the amount of digital health information grows each year, the challenge of effectively navigating through this vast array of content becomes increasingly daunting, raising concerns about misinformation and health-related misconceptions.

During the work on the comprehensive GAP (Gut informierte Kommunikation zwischen Arzt und Patient, meaning "Well-informed communication between physician and patient" in English) project, which addressed evidence-based information [18-26], we studied the needs and requirements of internet users for health information online [18]. We developed criteria to assess web domains according to these needs. Assessing the quality of websites is nothing new. There are existing seals of approval, such as the afgis (Aktionsforum Gesundheitsinformationssystem eV, meaning "Action Forum Health Information System eV" in English) and Health On the Net Foundation Code of Conduct seals, which check whether the website operator meets certain transparency and quality standards. An afgis certificate is valid for 1 year. However, with Health On the Net Foundation Code of Conduct, there were several issues [27], and on December 15, 2022, the Health On the Net Foundation discontinued its services. Different seals potentially signify different standards of certification. Another issue with these seals is that the pages that have a seal cannot easily be found via a central search. To address these issues by centrally and consistently rating domains inside a search platform, we developed a search engine in 2020. This search engine was originally called "GAP search" and later renamed as "tala-med Suche" in German and "tala-med search" in English.

The topic of data retrieval in medicine is generally relevant and widely researched. Efforts ranging from decentralized search systems for patient data in registries [28] to tools such as Informatics for Integrating Biology and the Bedside (i2b2) [29] that allow the creation of cohorts, for which data security plays an important role [30], have been investigated. In addition, for public data, such as the Medical Subject Headings (MeSH), search capabilities play an important role [31], including the requirement for an intuitive user interface design [32]. In terms of scraping web content, which could be used to create a search platform, the Sampled German Health Web [33,34] created an index of German-language health-related web pages using a special focused crawler. The Sampled German Health Web index was restricted to pages from the top-level domains .de, .at, and .ch, and the domains were automatically filtered for health-related content while crawling using a support vector machine. By contrast, the new tala-med search engine, which was used in this study, relied on hand-picked, high-quality, German-language health websites. These domains—more than 50 of them—underwent a rigorous evaluation process. This evaluation was based on the categories *trustworthiness* (with the subcategories *authority*, *independence*, and *evidence based*), *recency*, *user-friendliness*, and *comprehensibility*. The ratings of these categories influence the order of the results on the SERP. The search engine also values user privacy by avoiding data collection and advertising. The detailed design of the search engine is described in the *Methods* section and in Multimedia Appendices 1-3.

By assessing the acceptance and usability of this new technology, we aimed to improve the implementation of the search tool. Numerous studies in the field emphasize the significance of outcome measures as crucial indicators of a technology's effectiveness, user-friendliness, and potential for widespread adoption. User acceptance, often determined by perceived usefulness, ease of use, and individual attitudes, is critical for a technology's long-term viability [35,36]. Moreover, the usability of a technology, including learnability, efficiency, and user satisfaction, significantly affects user experience and reduces possible adoption barriers [37]. Evaluating these dimensions can provide insight into how well technology meets users' expectations, addresses their needs, and contributes to successful implementation. In the context of health care technology, this scrutiny is particularly relevant because effective tools for accessing medical information can significantly impact health care outcomes [17,38].

In this study, the acceptance and usability of our newly developed search engine was evaluated by means of a questionnaire study and actual use by tracking user behavior during the first months of operation.

Related Works

The usability and accessibility of digital platforms are essential for ensuring that all users, including those with disabilities, can effectively engage with websites and mobile apps. Mateus et al [39] conducted a systematic mapping of accessibility issues, revealing that automated tests covered <40% of accessibility problems on websites and even fewer on mobile apps. The study stressed the importance of including users with disabilities in evaluations because user testing uncovered many issues missed by automated tools and expert inspections. This underscores the need for a comprehensive evaluation approach that combines expert and real-world user input to improve accessibility.

Petrie and Bevan [40] expanded on this by exploring the interplay between usability, accessibility, and user experience. The authors defined usability in terms of effectiveness, efficiency, and satisfaction, focusing on aspects such as learnability, flexibility, and safety. Importantly, they emphasized that accessibility is an integral component of usability and that digital systems should cater to the widest possible range of users, including those with disabilities. They also introduced the concept of user experience, which goes beyond usability to include users' emotional responses and satisfaction with a system. This holistic approach is essential for creating digital platforms that are both functional and engaging for all users.

Belinda et al [41] provided a deeper look into the internal and external usability factors that affect website performance. The authors used automated tools such as GTmetrix and Website Grader to measure internal attributes such as performance, load time, and page size, while external attributes such as ease of navigation and user satisfaction were assessed through surveys. Their findings revealed that some websites performed well from a user perspective but were found lacking in technical performance, particularly in terms of load times and page requests. This highlights the need to address both internal and external usability factors to create well-rounded digital platforms.

Kritz et al [42] explored the online resources and tools used by European physicians to gather medical information. The authors found that physicians frequently relied on general search engines and faced significant barriers in accessing high-quality,

trustworthy medical content. Medical specialists were more likely to use medical research databases, while general practitioners often faced barriers such as lack of time and language restrictions. The study highlighted the need for improved medical search tools tailored to the specific needs of different physician subgroups. Kritz et al [42] concluded that user-centered medical search tools could significantly improve accessibility and the quality of online medical information.

Strecker et al [32] also contributed to the usability of medical search tools, focusing on the MeSH Browser. The authors evaluated a newly developed multilingual MeSH Browser, which introduced improvements in user interface design to enhance the accessibility and usability of medical literature searches. The results showed that contemporary web design principles led to significant improvements in navigation and overall user satisfaction, further underscoring the importance of continual evaluation and enhancement of information systems in the medical realm.

Eysenbach and Köhler [43] studied how consumers search for and appraise health information on the internet. The qualitative study revealed that users, despite using suboptimal search techniques, were able to retrieve health information quickly. However, users rarely checked critical indicators of credibility, such as the “About us” sections or disclaimers, relying instead on superficial factors such as professional design and ease of use. This reliance on superficial credibility indicators poses risks, particularly in the health care field, where the accuracy of information is paramount. Eysenbach and Köhler [43] suggested that further research is needed to develop educational and technological tools that guide users toward high-quality health information.

Zhang [44] expanded the understanding of how consumers select sources for health information by identifying 5 categories of factors that influence source selection: source-related factors, user-related factors, user-source relationships, characteristics of the problematic situation, and social influences. The study also identified a range of criteria that mediate the influence of these factors on source-selection decisions, including accessibility, quality, usability, interactivity, relevance, usefulness, and familiarity. Zhang [44] concluded that a personalized approach to health information systems is necessary to provide effective access to health information because different consumers prioritize different factors when selecting sources. This insight strongly indicates the need for more personalized information services that cater to individual user preferences and needs.

In conclusion, the literature emphasizes the critical role of accessibility, usability, and user experience in the design of digital platforms, especially health care websites. While Saad et al [45] have highlighted general usability problems in health care websites, the studies by Kritz et al [42], Eysenbach and Köhler [43], and Zhang [44] focus on the challenges that users face in retrieving and assessing the credibility of health information online. A personalized and user-centered approach to the design of health care information systems would improve the accessibility and quality of health information, meeting the diverse needs of consumers and health care professionals alike [40,44].

Methods

Search Engine Development

Requirements

The functional requirements for the tala-med search engine included the ability to crawl and index health information websites, implement quality assessment filters, and provide search term suggestions and synonym handling. The nonfunctional requirements focused on maintaining user privacy through self-hosting, ensuring a user-friendly interface, and optimizing performance to handle large sets of synonym mappings efficiently.

Implementation

Design

To implement tala-med search, we used software with a self-hosting capability because of privacy considerations. Furthermore, modern technology with single-page application design using a web application programming interface was required to compete with modern search engines. Therefore, the selection available was limited. As no single product met all our requirements, we built a search engine stack ourselves, consisting of crawler, middleware, and front end, using existing components and established technologies.

Back End

We created the web index using the open-source crawler software Fess [46]. For boilerplate removal, we used Mozilla’s Readability tool to strip HTML tags and display core content, cleaning up crawled websites for our search index [47]. As middleware, we used Elastic App Search [48] (now part of Elastic Enterprise Search [49]), which relies on Elasticsearch. This software allows for setting up weighting mechanisms for the search equation (Multimedia Appendix 1) and configuring filters. In our case, a set of criteria for quality assessment were developed, which we integrated as filters. These quality assessment criteria were adapted from a systematic review by Eysenbach et al [6] that identified criteria used from 1969 to 2001 for evaluating the quality of health information online. We identified 74 criteria [18] to assess the content quality of different health-related websites and rigorously evaluated >50 German-language health information providers (Multimedia Appendix 2) using these criteria, assessing 1 main page and 5 randomly selected subpages per provider. To compile the list of relevant providers, we started with a list from a Bertelsmann study [3] and adapted it with the help of domain experts. The 74 quality criteria were later condensed into 4 categories: *trustworthiness*, *recency*, *user-friendliness*, and *comprehensibility*. The scores of these categories were added to the search index after boilerplate removal and divided into 4 value ranges reflecting the quality of each category, using 4 quantiles for differentiation. These scores served as quality indicators and filter categories, influencing the order of the results on the SERP. More details about the internal mechanisms of the search ranking can be found in Multimedia Appendix 1.

App Search supports the consideration of synonyms, which is crucial in medical language due to varied etymology and numerous abbreviations. Our goal was to enable the search engine to handle synonyms effectively. In App Search, sets of up to 32 synonyms can be created, allowing synonyms of a matched word to be considered in searches. We generated synonym sets by collecting word types and pairs from the document text, filtering against a German stop word list, resulting in approximately 6 million unique entries. These entries were matched against an experimental German interface terminology—SCT-GIT—linked to Systematized Nomenclature of Medicine–Clinical Terms (SNOMED-CT) [50], yielding approximately 40,000 mappings to SNOMED-CT codes. Synonym links in SCT-GIT were then used to add synonyms for each matching string, finding at least 1 synonym for approximately 30,000 SCT-GIT terms. We limited the number of synonym sets to 500 to maintain performance, using only sets with at least 4 synonyms, resulting in a median set size of 4.

Front End

The front end was custom developed with a slim design inspired by Strecker et al [32], featuring search term suggestions and a user-friendly filter selection (Figure 1). After entering a search term, the interface displayed a SERP with evaluated German-language health information providers, showing their scores in the 4 filter categories through graphical indicators (Figure 2). Hovering over these graphics revealed explanations, and the bottom of the page included pagination and a footer with logos and links to pages displaying the imprint and data protection information.

Figure 1. Landing page of the search displaying the initial search query field.

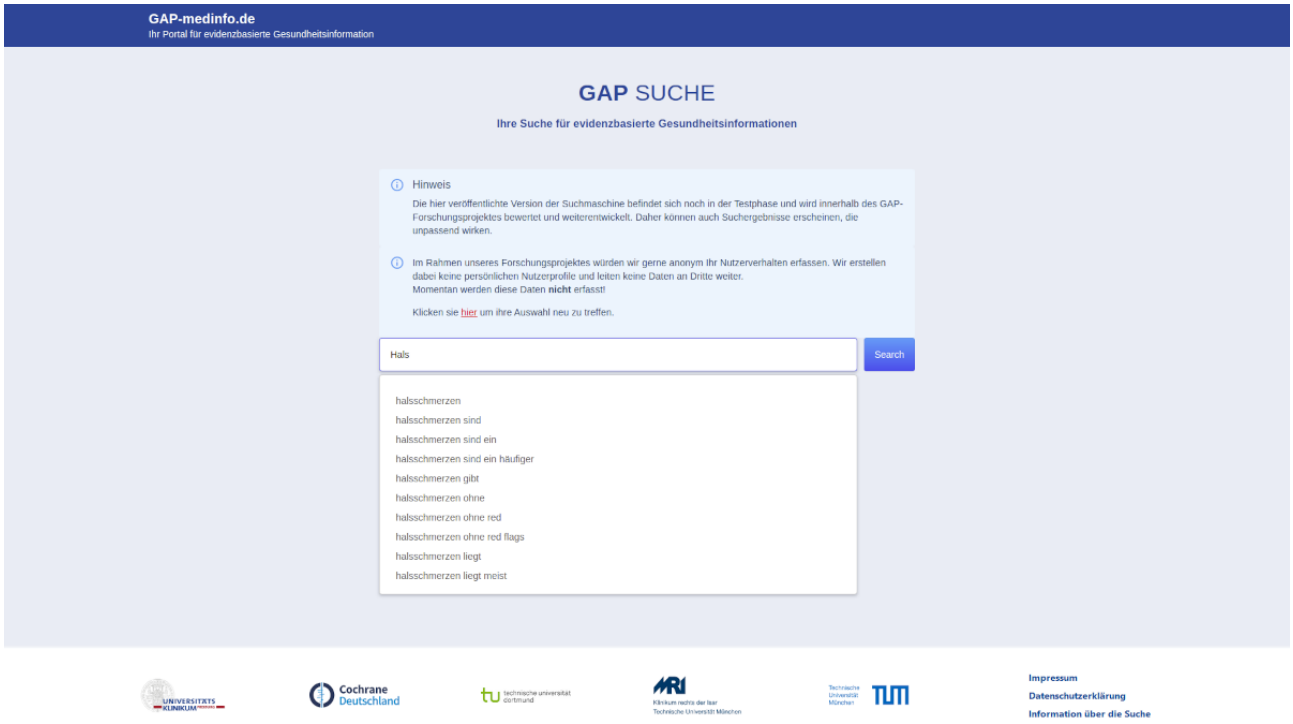
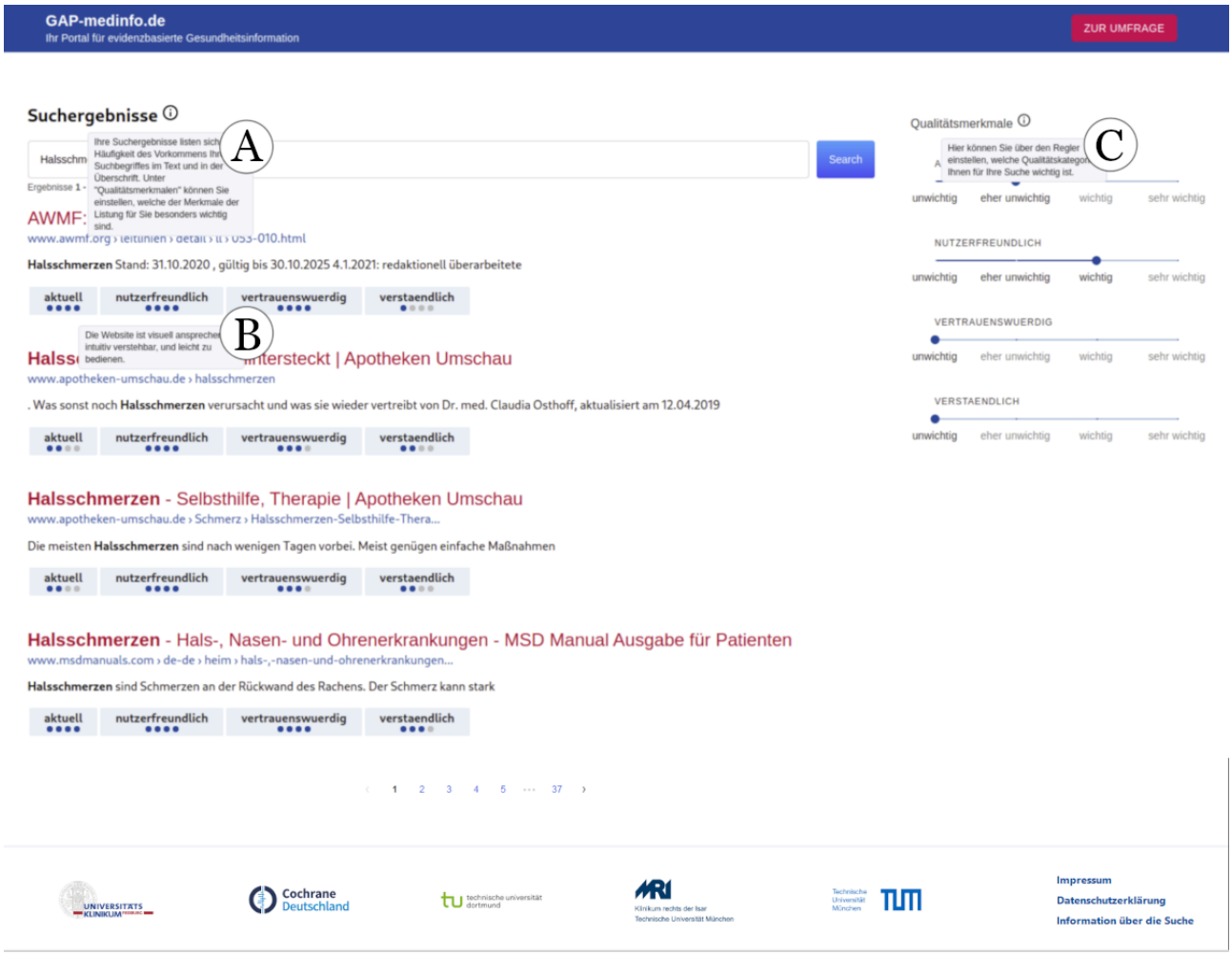


Figure 2. The search engine results page consists of 3 areas. (A) The search query. (B) List of search results. Each search result shows the quality in the 4 filter categories. (C) Slide controls allow users to set a threshold for each filter category (aktuell: recency, nutzerfreundlich: user-friendliness, vertrauenswürdig: trustworthiness, and verständlich: comprehensibility). In each of these areas, more information can be displayed by means of tooltips, as shown in the graphic (next to each label).



Pilot Testing

The search engine was pilot-tested between October 2020 and the end of June 2021. Adapting to the challenges posed by the COVID-19 pandemic, the pilot testing format shifted from a face-to-face setting to a web-based setting with a small group of academic staff from the University of Freiburg. The evaluation, involving 12 participants, consisted of screen recordings of search tasks and a qualitative survey regarding user-friendliness. After the pilot tests, technical enhancements were implemented, guided by the qualitative insights. These findings formed the basis for the subsequent questionnaire study.

Availability

The current version of the search engine can be accessed on the web [51]. Please note that there have been some significant changes to the website structure since this study was conducted.

Evaluation

Design

We used 2 data collection methods to evaluate user behavior. First, we asked users to fill out a questionnaire (self-perception). Second, we recorded user behavior via web tracking independently and unreferenced to the questionnaire (external perception). The study protocol was published in 2019 [22].

Questionnaire Study

The study used a cross-sectional design with data collected through a web-based questionnaire. The questionnaire was developed after pilot testing of the search engine and underwent an internal pretesting round before recruitment began.

Recruitment

The study used a recruitment strategy aimed at enrolling 200 participants. The first rounds of recruitment, starting in October 2020, which involved sending email invitations to local health initiatives, professional networks for physicians, and personal contacts, were unsuccessful in obtaining the desired number of participants. Participants were then recruited from the WisoPanel, an established online panel of German-speaking individuals [52]. The panel provided a pool of 14,900 potential respondents who had previously expressed an interest in participating in research studies. The survey was open for anyone accessing the link, but only participants from the panel (>70% of the data) were included in the analysis of this study to maintain a credible sample. Recruitment via the panel took place in May 2021. The survey ended at the end of May 2021.

Panel members were invited to participate in the study via email. The email informed participants about the project and the aspects distinguishing the new search engine from other platforms.

To ensure that participants had firsthand experience with the platform before proceeding to the survey, they were asked to test the search engine first. We recommended that participants carry out 1 or 2 searches and familiarize themselves with the SERP and the features and functions of the website.

Survey Administration

The hyperlink that led participants to the search engine was sent to the panel members by email. Subsequently, the link to the survey was displayed as a pop-up element after 10 to 20 seconds of interaction with the search engine. Throughout the assessment phase, a button to access the survey was strategically positioned in the top right corner of the website to allow a direct link to the survey interface. No registration was required to complete the survey, ensuring seamless and voluntary participation. Participants had the option to cancel their responses at any time, and there were no time restrictions regarding the completion of the questionnaire. Back buttons allowed participants to review their answers before submitting the survey. Conventional technical methods, such as cookies, were used to prevent visitors from submitting their survey more than once.

The web-based survey was conducted using the Unipark tool [53], with the survey page seamlessly embedded into the search page via a modal window. After inviting panel participants, the survey remained accessible for 11 days and closed at the end of May 2021.

Before filling out the survey, participants were provided with an introductory text outlining the anticipated time commitment, project objectives, details about anonymity, the option to terminate the survey at any point, data protection measures, and contact information for queries. To proceed, participants were required to confirm their understanding by checking a box and consenting to the specified use of their data (informed consent). No personal data were collected in the survey, and participants were instructed not to provide any identifying information in their open responses. Access to the data was restricted to members of the institutions participating in the study.

Survey Characteristics

On the basis of established scales (the German version of the self-assessment eHealth Literacy Scale [G-eHEALS] [54] and the System Usability Scale [55] adapted by Quirmbach [56] and Magin et al [57]), we developed a 25-item questionnaire across five dimensions: (1) sociodemographic data; (2) internet use; (3) digital health literacy; (4) usability, acceptance, and innovative aspects of the search engine; and (5) search filters. The final questionnaire contained 24 single-choice questions and 1 open-ended question (Tables S1 and S2 in Multimedia Appendix 4). The items in the questionnaire were not randomized. No form of adaptive questioning was used. All questions, except for the open-ended question, were mandatory; nonresponse options were not provided. Thus, there was no additional consistency or completeness check before submission. The survey took approximately 5 minutes to complete.

Statistical Analysis

Descriptive statistics were used to summarize the characteristics of the study participants, including age, gender, occupation, educational level, and internet use for health information, as well as self-perceived digital health literacy. Only completed questionnaires were analyzed. No statistical correction was applied to adjust the sample. To assess the acceptance and usability of the search engine, Likert-scale items related to these dimensions were combined to create scale values. The acceptance scale measured participants' willingness to use the search engine again and to recommend it to others. The usability scale evaluated participants' agreement with statements related to comprehensibility, clarity, the effectiveness of the search, confidentiality, advertising, the ease of learning, and using the search engine.

One reviewer analyzed the content of the open-ended question and assigned the comments to different categories inductively.

Multiple regression analyses were conducted to explore predictive factors influencing the acceptance of the search engine. The factors tested in the regression analyses included age, gender, occupation, educational level, the frequency of internet use for health information, and self-perceived digital health literacy.

Reporting

This study is reported based on the guidelines of the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [58].

User Behavior Tracking

Behavior tracking data were collected from all consenting users accessing the search engine between October 2020 and June 2021 using the web analytics software Matomo (Matomo.org) [59]. Matomo is a powerful open-source web analytics platform designed to help website owners and organizations gain insights into their online presence while respecting user privacy. Its primary purpose is to track website traffic by tracking website visitor behavior. What sets Matomo apart from services such as Google Analytics is its commitment to data privacy and General Data Protection Regulation (GDPR) compliance. Matomo allows users to maintain control over their data by hosting the data on their own servers, ensuring that sensitive information is not shared with third parties. Matomo is easy to integrate into web pages and can also be customized to track specific custom data. Ripp and Falke [60] successfully used Matomo to track user behavior in their research project, where they analyzed search behavior for certain keywords on the online information system Grammis.

We installed Matomo on our own server and integrated the tracking into our search engine's front end. We enabled the tracking to record the filter settings and capture the search term beyond the standard visit data, such as how long users spent on the website, and which of the results they clicked on. Structurally, Matomo records several actions for each visit, containing the duration among other data (Figure S1 in Multimedia Appendix 5). However, for the last action of a visit, the duration is not recorded. We have filled the empty value

with the median of all actions from the particular visit to obtain a more realistic value for the entire visit duration.

Only data from consenting users was recorded and subsequently descriptively analyzed for the 9-month time period. Therefore, this sample differs from the sample of the questionnaire study but may also include users who completed the questionnaire. The data processing for the data analysis was realized with a Python script directly accessing Matomo's database using specific SQL queries.

Ethical Considerations

The study was approved by the institutional review board at the University of Freiburg (Ethikkomission-Freiburg 559/17) as an extension to the GAP study [22], and a document on data protection was agreed upon. As an incentive, the questionnaire study participants could take part in a prize draw for 1 of 25 vouchers to a bookstore worth €20 (US $24.33) after completing the survey.

Results

Questionnaire Study

Participants

During the time the survey was open (from October 2020 to the end of May 2021), there were 1577 unique site visitors to the search engine website, with 1426 (90.4%) visitors to the first survey page. Of these 1426 visitors, 1250 (87.7%) agreed to participate in the survey. Of these 1250 participants, 1123 (89.8%) completed the survey.

Of the 1123 survey respondents, 802 (71.4%) were from the panel described in the Recruitment subsection. The age distribution of these participants was approximately normal, with a little more than a quarter of the participants (215/802, 26.8%) falling into the age category of 50-59 years. The gender distribution was also fairly balanced between male and female, with a little more than half of the respondents (432/802, 53.9%) self-identifying as female. Regarding occupational background, the majority of the participants (726/802, 90.5%) did not work in a medical profession.

Educational levels were notably higher than in the general population due to recruitment through the scientific panel, with 69.3% (556/802) of the participants having completed upper secondary education (at least level 3 in the International Standard Classification of Education [ISCED]-2011) and 46.3% (372/802) holding a bachelor's or master's degree or higher (ISCED-2011 levels 5-8).

The study participants recruited from the web-based panel reported searching for health information online more frequently than the general population, with 65.4% (524/802) searching several times a month at the minimum and 27.7% (222/802) searching at least several times a week. Only 1.5% (12/802) of the participants indicated that they "never" searched the internet for health information.

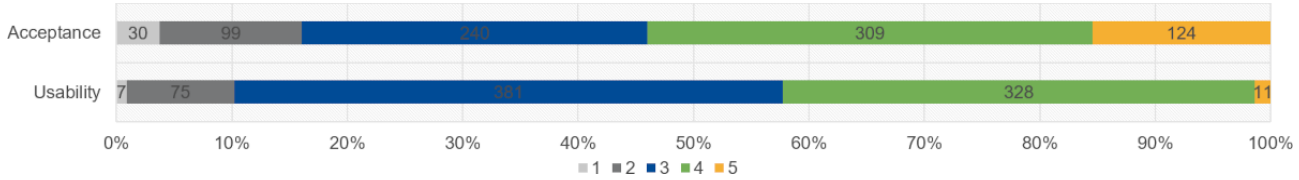
Participants' self-perceived digital health literacy was measured by the G-eHEALS items. Of the 802 participants, 291 (36.2%) felt confident in making health-related decisions based on internet information, while 461 (57.5%) felt capable of distinguishing reliable from questionable information online.

More information about the participants is available in Table S3 in Multimedia Appendix 4.

Acceptance and Usability

Scale values were created by combining items relating to the acceptability and usability of the search engine (Figure 3).

Figure 3. Scale values for acceptance and usability (n=802). The Likert-scale response options were as follows: 1=strongly disagree, 2=disagree, 3=partially agree, 4=agree, and 5=strongly agree. There was one negative question, which was adjusted for calculating the scale values.



The results show agreement regarding the acceptance and usability measures of the new search engine (Figure 3 and Table 1). In particular, more than half of the participants expressed their willingness to use the search engine again (465/802, 58%) and to recommend it to others (507/802, 63.2%). In addition, 79.3% (636/802) of the respondents agreed or strongly agreed that using the search engine was quick to learn, and 77.3% (620/802) to 98.6% (791/802) agreed at least partially with other statements related to usability aspects.

Table 1. Detailed items for acceptance, usability, and innovative aspects (n=802). The Likert-scale response options for acceptance and usability were as follows: 1=strongly disagree, 2=disagree, 3=partially agree, 4=agree, and 5=strongly agree. The Likert-scale response options for the innovative aspects item were as follows: 1=very unimportant, 2=unimportant, 3=neither unimportant nor important, 4=important, and 5=very important.

Content	Scores, mean (SD)
Acceptance	
Item 1: potential reuse	3.58 (1.02)
Item 2: recommendation to others	3.68 (1.03)
Usability	
Item 1: quick to learn	4.03 (0.86)
Item 2: efficient scannability of the SERP[a]	3.63 (1.00)
Item 3: clarity of the SERP	3.75 (1.00)
Item 4: functions *not* comprehensible[b]	2.37 (1.29)
Item 5: appreciation for the absence of advertising	4.66 (0.66)
Item 6: functionality meets expectations	3.45 (1.03)
Item 7: careful handling of personal data	3.59 (0.88)
Item 8: no commercial bias	3.55 (0.95)
Item 9: fast access to relevant information	3.77 (0.99)
Item 10: innovative approach	3.58 (1.03)
Innovative aspects of the search engine (anonymous searches)	4.50 (0.76)

[a]SERP: search engine results page.

[b]Item 4 (comprehensibility of functions) was a negated question.

Users particularly appreciated the absence of advertising on the platform, with 93.1% (747/802) agreeing or strongly agreeing that the absence of advertising was a significant benefit. The search engine's user-friendly interface and the ability to adjust search results through the use of filters were also well received by participants. Furthermore, the unique features of the search engine, notably the anonymous searches without the creation of user profiles and independence from sponsors, received high importance ratings from 91.3% (733/802) of those surveyed.

More detailed results are available in Tables S4-S6 in Multimedia Appendix 4.

Of the 802 participants, 561 (70%) to 622 (77.6%) indicated that they used the filter functions that could be enabled to change what was displayed on the SERP (Table 2; Figure 2). Furthermore, 92.3% (518/561), 93.9% (553/589), 94.7% (567/599), and 96.5% (600/622) of those users who used the filters agreed at least partially that the filter functions were helpful in finding trustworthy, recent, user-friendly, and comprehensible results, respectively. The *recency* filter was particularly well received, with 46% (286/622) of the users strongly agreeing that it was helpful. More detailed results are available in Table S7 in Multimedia Appendix 4.

Table 2. Helpfulness of filters (n=802). The Likert-scale response options were as follows: 1=strongly disagree, 2=disagree, 3=partially agree, 4=agree, and 5=strongly agree.

Content	Participants, n (%)[a]	Values, mean (SD)
Helpfulness of filters for each filter		
Item 1: trustworthiness	599 (74.7)	4.11 (0.92)
Item 2: recency	622 (77.6)	4.25 (0.84)
Item 3: comprehensibility	589 (73.4)	4.05 (0.93)
Item 4: user-friendliness	561 (70)	3.91 (0.93)

[a]All other participants answered, "I did not use this function, I cannot evaluate this question."

In the open-ended question, which aimed to supplement the data with qualitative results, participants were asked to share their comments and suggestions for improvements. Of the 802 participants, 263 (32.8%) replied to the question. Multiple aspects were mentioned. In an inductive approach, 10 main themes were identified (Figure 4). Many of the participants commented on the usability aspects of the website. Of the 236 comments, 76 (32.2%) included constructive criticism and specific suggestions for improving the website in the future. Many commenters made suggestions for improving the design because they found some aspects unclear or overwhelming. Others did not fully understand the methodology behind the

evaluation of the selected providers or the reason why only specific providers, including commercial web pages, were shown on the results page and asked for more clarification. Some commenters were happy with the site as it stood and did not ask for any changes, while some conveyed disappointment with the results they discovered on the SERP.

Figure 4. In the analysis of 263 comments in response to the open-ended question, 10 main themes were identified inductively: design; search results and filter functions; selection of providers; technical details; method of evaluating the website; target audience; text on the website; comments about the survey; general positive responses; and critique, dissatisfaction, and other comments. The illustration shows a hierarchical visualization, with the size of the elements indicating the quantity of comments.

Predictors of Acceptance

We conducted multiple regression analyses to examine the predictive factors influencing responses to the acceptance scale. Age, gender, occupation, educational level, frequency of internet use for health information, and self-perceived digital health literacy were tested as influencing factors. All factors were first tested individually and subsequently tested simultaneously due to the potential error introduced by multiple testing. Overall, only minor discrepancies were identified. Significance levels were set at $P<.05$. Table 3 presents more details.

Table 3. Predictive factors for acceptance (n=802). Categorical variables were dummy coded.

Predictive factors for acceptance	Individual testing of each predictor				Simultaneous testing of all predictors[a]			
	B (SE)	*F* test (*df*)	*P* value	R^2 (%)	B (SE)	*F* test (*df*)	*P* value	η_p^2 (%)
Age	*0.07*[b] *(0.02)*	*7.97 (1)*	*.005*	*0.99*	*0.07 (0.03)*	*7.27 (1)*	*.007*	*0.91*
Gender		0.04 (2)	.96	0.01		0.20 (2)	.82	0
Diverse	0.01 (0.49)	0.00 (1)	.99		−0.07 (0.49)	0.02 (1)	.89	
Female	0.02 (0.07)	0.09 (1)	.77		0.04 (0.07)	0.35 (1)	.55	
Male	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Occupation		1.82 (2)	.16	0.45		1.14 (2)	.32	0.29
Medical doctor	0.12 (0.12)	0.92 (1)	.34		0.07 (0.12)	0.28 (1)	.59	
Other medical profession	−0.56 (0.35)	2.62 (1)	.11		−0.49 (0.35)	1.93 (1)	.17	
Not working in a medical profession	Reference	Reference	Reference	Reference	Reference	Reference	Reference	Reference
Highest level of educational attainment		*2.57 (4)*	*.04*	*1.27*		1.55 (4)	.19	0.78
No formal secondary education (ISCED[c] levels 0-1) or lower secondary education (*Hauptschulabschluss*, ISCED-2011 level 2)	*0.51 (0.19)*	*6.81 (1)*	.009		*0.41 (0.20)*	*4.22 (1)*	.04	
Lower secondary education (*Realschulabschluss*, ISCED-2011 level 2)	0.33 (0.17)	3.85 (1)	.05		0.26 (0.17)	2.33 (1)	.13	
Upper secondary education (*Fachabitur oder Abitur*, ISCED-2011 levels 3-4)	0.16 (0.17)	0.94 (1)	.33		0.13 (0.17)	0.58 (1)	.45	
Tertiary education: bachelor's or master's degree or equivalent (ISCED-2011 levels 6-7)	0.21 (0.16)	1.67 (1)	.20		0.17 (0.16)	1.14 (1)	.29	
Doctorate degree (ISCED-2011 level 8)	Reference	Reference	Reference	Reference	Reference	Reference		Reference
Frequency of internet use for health information	*0.15 (0.04)*	*17.16 (1)*	*< .001*	*2.1*	*0.13 (0.04)*	*11.98 (1)*	*<.001*	*1.5*
Self-perceived digital health literacy: information search	*0.12 (0.04)*	*9.69 (1)*	*.002*	*1.2*	0.08 (0.05)	3.03 (1)	.08	0.38
Self-perceived digital health literacy: information assessment	0.05 (0.04)	1.79 (1)	.18	0.22	−0.01 (0.05)	0.04 (1)	.85	0

[a]Variance resolution of simultaneous testing: R^2=5.02%.

[b]Factors with significant predictions (*P*<.05) are shown in italics.

[c]ISCED: International Standard Classification of Education.

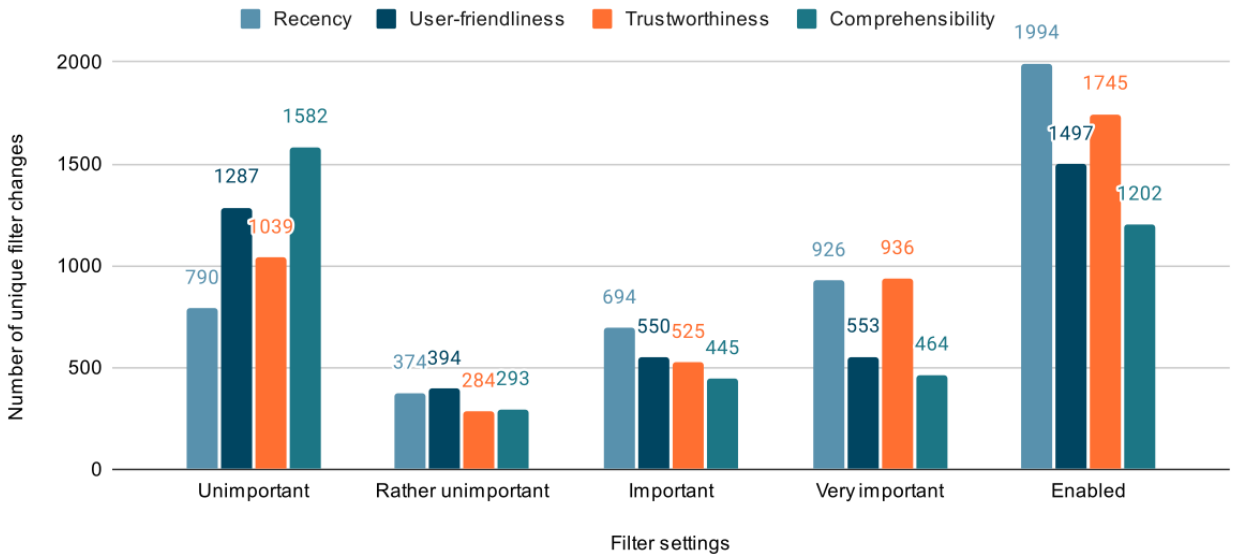
There were 2 significant results in predicting the acceptance of the search engine. First, the acceptance rating increased with age, rising by 0.07 scale points (B=0.07) for each additional 10-year age group. Second, the ratings for acceptance were higher among participants with more frequent internet use for health information.

The factors "highest level of educational attainment" and "self-perceived digital health literacy" only showed significant correlations with the acceptance ratings in the individual tests and slightly lower correlations in the simultaneous tests. Participants with lower levels of education showed stronger acceptance ratings than those with higher levels of education. For this predictor, the reference category was set as "doctorate degree" (ISCED-2011 level 8). Acceptance ratings among persons without any formal secondary education (ISCED-2011 levels 0-1) or with a lower secondary education (*Hauptschulabschluss*, ISCED-2011 level 2) were 0.41 scale points higher than among persons with a doctorate degree (B=0.41). However, if we examine the effect strength using partial eta–squared (η_p^2), only small effects were shown overall for all factors (1%=small effect, 6%=medium effect). No significant results were found for the factors *gender* or *occupation*.

User Behavior Tracking

For a period of 9 months (from October 2020 to the end of June 2021), user behavior of consenting users was tracked using Matomo. The main aspects recorded were the search queries and the use of the filter functions. The data indicated that 1631 visitor sessions had taken place, with 3090 searches using 1984 different search terms across 1924 visits. The search terms had an average lexical token count of 1.74 (SD 1.14). Of the 1984 search terms, 1096 (55.24%) had a lexical token count of 1, while 888 (44.76%) had a lexical token count of >1. The maximum lexical token count was 13. We observed that 28.9% (893/3090) of the search queries were conducted through mobile devices such as smartphones and tablets. On average, visitors conducted 1.64 (SD 1.31) searches per visit, with an average duration of 137.45 (SD 278) seconds and a median duration of 49 (IQR 24-119) seconds. The top 3 search terms were "corona," "rückenschmerzen" (back pain), and "husten" (cough; Table S1 in Multimedia Appendix 5). Each search term was queried 1.57 (SD 3.46) times on average; of the 1984 search terms, 1621 (81.7%) were entered just once. Users could personalize the order of the displayed search results, depending on the prioritization of the different filter categories (*trustworthiness*, *recency*, *user-friendliness*, and *comprehensibility*), all of which were initially set to "unimportant" by default. Within the search interactions, overall, 47,532 actions were performed, among which were 1358 (2.86%) outlink actions (ie, external domains clicked by website visitors) and 34,490 (72.56%) filter changes in total. The high number of filter changes was due to the fact that the filters were often switched back and forth in a search, which led to multiple occurrences of the same filter settings. This behavior was probably observable due to the fact that, given the limited corpus, the result set could be empty for certain search terms in combination with a narrow filter setup; therefore, it was necessary to change back the filter to obtain results. Excluding these repeated changes, of the 34,490 filter changes, 5534 (16.05%) were unique. In addition, after removing the initial filter setup where all categories were set to the default value, we ended up with 2784 nondefault unique filter changes used in 15.37% (475/3090) of the searches. *Recency* was used the most, followed by *trustworthiness*, *user-friendliness*, and *comprehensibility* (Figure 5). Consistent with the results on the usefulness of the filter function from the questionnaire study, the user behavior data showed that the filters for *recency* and *trustworthiness* were adjusted most frequently.

Figure 5. Distribution of the filter settings of the 2784 nondefault unique filter changes across filter categories. We consider a filter category to be "enabled" if a filter category is not set to "unimportant.".



Discussion

Principal Findings

In this study, we evaluated the independent, noncommercial tala-med search for German-language medical information, placing a primary focus on transparency, independence, and evidence-based information. Our questionnaire study and user behavior tracking provide insights into the acceptance and usability of this innovative platform.

The survey resulted in a mean acceptance score of 3.63 (SD 0.98) and a mean usability score of 3.76 (SD 0.61) on a scale of 1 to 5 (1 signifying strong disagreement and 5 signifying strong agreement with acceptance and usability statements). Given that our user interface is similar to that of popular search engines such as Google and Bing and is advertisement-free, it is unsurprising that it was well received. However, there is still a need for optimization to achieve a higher reuse potential and a greater willingness to recommend the platform. Within the usability measures, item 1, which reflected the learnability of the search, achieved a high mean value of 4.03 (SD 0.86; Table 1). In terms of the effectiveness and clarity of our design (refer to items 2, 3, and 9 in Table 1), we still see room for improvement but are satisfied with the results for the initial release. Our findings align with those of Petrie and Bevan [40], who emphasized that usability, particularly learnability, is crucial for ensuring the platform's effectiveness and user satisfaction. Similar to the study by Petrie and Bevan [40], our study highlights the need to continually evaluate and refine

digital systems to cater to user needs and improve overall usability. The survey also revealed that the filters were used by >70% of the 802 users. Among these users, the filters were highly accepted on average, which matches with the tracking statistics revealing the high use of the filter functionality. The most used filter was *recency*, followed by *trustworthiness*, matching the survey outcome, which asked about the helpfulness of the filters. This was not the case for the 2 remaining filters. Although the filter for *user-friendliness* was used more often in our tracking data than the one for *comprehensibility*, it was perceived as less helpful than comprehensibility. In the future, we might inspect these filter categories in more detail and potentially improve their estimation. The insights support our decisions, including basing the design of our search engine on the insights of Strecker et al [32], which demonstrated how improvements in user interface design could lead to higher user satisfaction. The high acceptance of the filters for *recency* and *trustworthiness* in our study also echoes the findings of Strecker et al [32], who highlighted that content navigation and relevance contribute to overall user satisfaction. Nevertheless, the insights encourage us to improve and possibly expand the filters and the design. The fact that our search engine does not share any data with partners or advertisers and does not record the visitor's behavior to create profiles was rated as very relevant, with a mean score of 4.5 (SD 0.76). This result confirms the importance of our general endeavor driven by data privacy and noncommerciality as our fundamental motivations for creating the search engine. This finding is consistent with the work of Eysenbach and Köhler [43], who found that while internet users claimed to prioritize credibility when assessing health information online, they often relied on superficial factors such as website design and ease of use, rather than more reliable indicators such as the "About us" sections or disclaimers. Our platform addresses these shortcomings by offering more explicit filters, such as *trustworthiness*, which help users find credible information without having to rely solely on superficial design elements. In our calculations on the predictors of acceptance, it can be concluded that while the effects may be small, the search engine was particularly well accepted among older users, users with a high frequency of internet use for health information, and those with lower levels of education. This finding suggests that the new search engine may effectively cater to segments of the population with lower health literacy and digital health literacy and greater requirements for navigational assistance, bridging the gap in accessing reliable medical information online. As discussed by Kritz et al [42], general practitioners and medical specialists often face barriers in accessing high-quality medical content due to limited time and the overwhelming amount of information online. Our search engine's focus on evidence-based information and user-friendly design helps mitigate these issues, making it a suitable tool for both health care professionals and the general public.

This reinforces the necessity to consider navigational needs in health information platforms. Digital health literacy has improved in recent years, particularly among persons with low levels of education [14]. The ability to assess health information has also improved over time and over the course of the COVID-19 pandemic. Digital health literacy improved for younger but not for older people over the course of the pandemic [15]. Even with some of these improvements, health literacy is still low within the German population and particularly among people with low levels of education. This observation aligns with the findings by Zhang [44], who identified that source selection for health information is influenced by a wide range of factors, including accessibility, quality, usability, and personal relevance. Our platform's appeal to older users and those with lower health literacy supports the conclusion drawn by Zhang [44] that a personalized approach, considering different user-related factors, is crucial for effectively meeting the diverse needs of consumers when they access health information. The correlation between low educational levels, low social status, higher age, and lower health literacy rates has become even stronger [14]. This emphasizes the need to address low health literacy to improve social inequities. The new search engine attempts to address this need by assessing health information online and by making high-quality information easily searchable.

Many of the comments highlight the fact that at the time of the survey, participants used the first version of the search engine, which still had some technical and appearance-related weaknesses. Since the survey was conducted, a number of revisions have been made, some of which correspond to suggestions made in the comments, such as adding more information about the principles of the quality assessment and the functions of the websites and removing certain commercially funded information sources.

The platform's emphasis on anonymous searches and independence from advertising and sponsors resonated especially well with the participants, underscoring the importance of developing alternatives to commercial search platforms. The usability of our platform, free from external advertising influence, aligns with the findings by Belinda et al [41], who evaluated both internal factors (such as page load time and performance) and external factors (such as ease of navigation and organization of information) as critical to the overall user satisfaction with websites. Our platform similarly emphasizes ease of use and efficient performance, which were key contributors to its positive reception. The general results align with previous findings on the importance of user-friendly interfaces [61-63] and the impact of advertising on users' perceptions [64].

Limitations

Limitations of the Technology

Some participants expressed dissatisfaction with the displayed results. Several respondents found the design of the SERP and individual results confusing due to an overload of information in one place. Potential design changes have been suggested to improve clarity. Currently, some search results are duplicated on the results page. To enhance the appearance of the SERP, it is desirable to deduplicate the search index. This would mean displaying more relevant information and fewer distracting results. To address the fact that sometimes no results were displayed due to strict filter settings, we could display results with nonexact matches for the filter category score below the exact matches and visually indicate them.

Participants encountered difficulties in understanding how the evaluation process worked. Efforts should be made to make the evaluation process easier to understand when using the website. Some improvements have been made over time, such as providing brief explanations of the filter categories when hovering over the graphic showing the score for each domain.

In addition, further clarification is necessary regarding the domains chosen, which comprise both independent and commercial providers. Over time, certain websites were excluded from being displayed on the site because of the conflicts of interest of their publishers (such as the site "Zentrum der Gesundheit" and a site funded by a pharmaceutical company). Currently, the website no longer displays commercial providers.

Limitations of the Questionnaire Study

While the questionnaire study provides valuable insights into the acceptance and usability of the new search engine, several limitations should be acknowledged. First, the recruitment methodology, relying on web-based panel recruitment, unintentionally led to an overrepresentation of individuals with higher educational levels than the general population and potentially more frequent internet use due to the online recruitment process. Among our participants, 46.3% (371/802) had completed tertiary education compared to only 18.5% of the general German population with a university degree or equivalent [65]. Similarly, 65.4% (524/802) of our participants reported searching for health information only at least once a month, while a Bertelsmann Foundation 2018 study indicated that only approximately 50% of the general German population seek such information at least once a month [2]. This potentially influences acceptance and usability ratings. Participants' self-perceived digital health literacy, as measured by the G-eHEALS items, approximately aligned with that of the sample of the validation study of the German version of the questionnaire by Söllner et al [54]. However, this information should be viewed with caution because Kim et al [66] discovered a discrepancy between self-assessments and actual ability. Actual ability was not assessed in our study.

Second, as some of the participants only interacted with the search engine and its features for a brief duration, this short testing time proved to be a limiting factor, hindering the formation of comprehensive first impressions. The responses to the open-ended question suggest that some participants tested the features of the search engine for only a few seconds. This was caused by the pop-up window linking to the questionnaire opening after interacting with the website for 10 to 20 seconds. Such a brief interaction period was likely insufficient for users to thoroughly explore and understand the functionalities and benefits of the search engine. Consequently, the feedback provided by users on the acceptance and usability of the platform may not accurately reflect its true potential. The limited duration of testing may lead to superficial evaluations, whereby participants may base their judgments on initial impressions rather than informed use. This can result in an inaccurate assessment and therefore impact the results presented for the acceptance and usability measures. More extended testing periods could offer a more accurate assessment of the platform's acceptance and usability over time. Furthermore, the implementation of controls for the length of the testing period could provide more reliable data, thereby ensuring that all participants have sufficient time to engage with the platform before completing the questionnaire.

Limitations of the User Behavior Tracking

As Matomo does not record the last action, we decided to substitute this missing value with the median of all actions from the particular visit. However, a more optimal solution would be to record of the last action, which could be triggered when closing the browser, but Matomo does not support this feature.

Unfortunately, due to data protection regulations, we were unable to link the search behavior with the survey data. It would have been interesting to see to what extent the search behavior and the questionnaire would have provided more information.

Future Work

Further Evaluation of the New Search Engine

As these results are based on the first impressions of a group of users, data on long-term use and satisfaction have yet to be compiled. Future work should therefore address the need to evaluate long-term acceptance and usability measures of the site using a longitudinal study design. This could provide valuable insight into whether users actually use the search engine when they have an acute need to seek out medical information and whether they return to the site over time.

Future research should also consider comparing the search engine with other existing and established platforms to gain a deeper understanding of its unique advantages. Such investigations might involve participants executing identical search queries on both the novel search engine and other platforms and sharing their impressions of each experience, similar to the study by Strecker et al [32]. Particular attention should be given to participants' abilities to assess the quality and reliability of the information found. Furthermore, a comparison of the specific results retrieved from both search engines could be undertaken, examining both the congruence and discrepancies in the results.

Further Improvements to the Search Engine

Ongoing revisions and updates to the search engine, including integration on partner websites, hold promise in expanding its reach and use, ensuring that it remains relevant and effective in addressing users' health information needs. Since the data collection for this study was completed, there have been several revisions to the website.

As the qualitative survey showed, some users found the SERP complicated. To address this issue, a potential solution could be a clearly linked help page with a video explanation. Another solution could be to provide an animated assistant that guides users through the page.

One major disadvantage of the search engine in its current state stems from the fact that the evaluation of the domains—more than 50 of them—according to the quality criteria was performed in 2020, and there has not been any update to the evaluations thus far. The domains' quality may have improved or declined

in the time since the evaluation. A continual process with constant reevaluations of the domains is necessary to ensure an up-to-date rating of the domains accessible through the search engine. As we currently do not have the capacity to perform another evaluation or to establish a continual evaluation process for the domains, we have temporarily disabled the display of the filter category scores and the option to adjust the results based on these scores because the scores do not reflect the actual quality of the domains in their current state.

Thus, in future, inspired by Zowalla et al [34], we intend to evaluate various criteria automatically. In the study by Zowalla et al [34], the readability of the language was analyzed supported by a support vector machine. We plan to use more sophisticated approaches that also model contextual context. Therefore, we plan to create classifiers based on Bidirectional Encoder Representations from Transformers (BERT) [67]. We have in mind the domain-specific models GerMedBert [68] and BioGottBERT [69], the latter being a model of the German monolingual Robustly Optimized BERT Pretraining Approach (RoBERTa) [70] model GottBERT [71] specialized on medical language. A further improvement might be the enrichment of the texts with MeSH terms, similar to PubMed. Such approaches would enable us to process and rate documents individually, rather than evaluating domains using random samples.

From a technological standpoint, there are still notable areas for enhancement of the middleware, which could be implemented in the future. As App Search is proprietary and monolithic, to make it customizable and more extensible, we are currently exploring potential improvements. In particular, we are considering alternative technologies, as used by Scheible et al [31]. This would enable us to add features based on modern technologies. Specifically, we are contemplating a different approach to computing synonyms. Instead of using discrete structures, we plan to train and use a FastText model [72].

Conclusions

The development of the independent, noncommercial tala-med search marks a step toward improving access to reliable and evidence-based health information online. By prioritizing transparency, independence, and high-quality content, the platform has the potential to bridge the gap in health information accessibility, especially for older and less-educated individuals. This study achieved its objective of evaluating the acceptance and usability of the platform, using user questionnaires and behavior tracking, and the results have been encouraging. The search engine was well received, with users appreciating its advertising-free design and filtering system, which addresses the navigational needs of individuals with lower health and digital health literacy.

User feedback also highlighted areas for improvement, particularly in the clarity of the design and filter functionality. Nevertheless, the platform's emphasis on anonymous searches and independence from advertising resonated strongly with participants, reinforcing the importance of alternatives to commercial search engines. These findings are based on initial impressions, with long-term use and satisfaction data still to be collected. Future research should compare the search engine with existing platforms to further explore its unique advantages, while ongoing updates and potential integration on partner websites will enhance its reach and relevance.

In conclusion, the tala-med search demonstrates a promising step toward enhancing health literacy and empowering individuals to make well-informed health decisions. By addressing users' navigational needs and promoting equitable access to high-quality, evidence-based medical information, the search engine has the potential to positively impact health literacy, reduce health disparities, and promote patient empowerment in the digital age.

Acknowledgments

The study was funded by the Innovationsausschuss of the Gemeinsamer Bundesausschuss (01NVF17010) as part of the GAP (Gut informierte Kommunikation zwischen Arzt und Patient, meaning "Well-informed communication between physician and patient" in English) project. Furthermore, this work was supported by the German Ministry for Education and Research (01ZZ1801B, 01ZZ1804A, 01KX2121, and 01ZZ2304A). Finally, we acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Klaus Koch of the Institute for Quality and Efficiency in Health Care supported the domain selection for the search index. Julia Serong and Jakob Henke of the TU Dortmund University contributed to the evaluation concept.

Authors' Contributions

RS, NK, and CW planned and carried out the implementation of the search engine supported by FT under the supervision of MB and SV-R. RS carried out the mathematical framework for the search engine and filter estimation with the help of LS. KW and AS-L worked out the criteria with which the domains were evaluated. MS created the procedure for the weighting of the filter categories and evaluated the pages. MS developed the questionnaire under the supervision of AM. CS prepared the data analyses for the predictors of acceptance under the supervision of EF-G. CS and RS prepared the data analyses of the web tracking. RS and LS wrote the manuscript with the help of all authors and under the supervision of AM. The manuscript has been read and approved by all named authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search engine development: search equation and filter estimation.
[PDF File (Adobe PDF File), 830 KB - humanfactors_v12i1e56941_app1.pdf]

Multimedia Appendix 2
Search engine development: list of domains.
[PDF File (Adobe PDF File), 105 KB - humanfactors_v12i1e56941_app2.pdf]

Multimedia Appendix 3
Search engine development: criteria evaluated.
[XLSX File (Microsoft Excel File), 19 KB - humanfactors_v12i1e56941_app3.xlsx]

Multimedia Appendix 4
Questionnaire study: questionnaire and detailed results.
[PDF File (Adobe PDF File), 313 KB - humanfactors_v12i1e56941_app4.pdf]

Multimedia Appendix 5
User behavior tracking details.
[PDF File (Adobe PDF File), 156 KB - humanfactors_v12i1e56941_app5.pdf]

References

1. Baumann E, Czerwinski F, Rosset M, Seelig M, Suhr R. [How do people in Germany seek health information? Insights from the first wave of HINTS Germany]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Sep;63(9):1151-1160. [doi: 10.1007/s00103-020-03192-x] [Medline: 32666180]
2. Haschke C, Westrick MG, Schwenk U. SPOTLIGHT gesundheit: gesundheitsinfos. Bertelsmann Stiftung. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/spotlight-gesundheit-gesundheitsinfos [accessed 2022-05-16]
3. Marstedt PG. Das internet: auch ihr ratgeber für gesundheitsfragen? Bertelsmann Stiftung. 2018 Jan 26. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/das-internet-auch-ihr-ratgeber-fuer-gesundheitsfragen [accessed 2022-05-18]
4. Individuals - internet activities. Eurostat. URL: https://ec.europa.eu/eurostat/databrowser/view/isoc_ci_ac_i/default/table?lang=en [accessed 2024-11-26]
5. Zhang Y, Sun Y, Xie B. Quality of health information for consumers on the web: a systematic review of indicators, criteria, tools, and evaluation results. J Assoc Inf Sci Technol 2015 Oct;66(10):2071-2084. [doi: 10.1002/asi.23311]
6. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
7. Hambrock U. Die suche nach gesundheitsinformationen. Bertelsmann Stiftung. 2018 Jan 26. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/die-suche-nach-gesundheitsinformationen [accessed 2022-05-18]
8. Baumann E, Czerwinski F. Erst mal doktor Google fragen? Nutzung neuer medien zur information und zum austausch über gesundheitsthemen. In: Böcken J, Braun B, Meierjürgen R, editors. Gesundheitsmonitor 2015: Bürgerorientierung im Gesundheitswesen. Gütersloh, Germany: Bertelsmann Stiftung; 2015:57-79.
9. Kammerer Y, Gerjets P. Effects of search interface and internet-specific epistemic beliefs on source evaluations during web search for medical information: an eye-tracking study. Behav Inf Technol 2012 Jan;31(1):83-97. [doi: 10.1080/0144929x.2011.599040]
10. Haas A, Unkel J. Ranking versus reputation: perception and effects of search result credibility. Behav Inf Technol 2017 Sep 26;36(12):1285-1298. [doi: 10.1080/0144929x.2017.1381166]
11. Cai HC, King LE, Dwyer JT. Using the Google™ search engine for health information: is there a problem? Case study: supplements for cancer. Curr Dev Nutr 2021 Feb;5(2):nzab002 [FREE Full text] [doi: 10.1093/cdn/nzab002] [Medline: 33937613]
12. Schaeffer D, Vogt D, Berens EM, Hurrelmann K. Gesundheitskompetenz der bevölkerung in Deutschland: ergebnisbericht. Universität Bielefeld, Fakultät für Gesundheitswissenschaften. 2017. URL: https://pub.uni-bielefeld.de/record/2908845#fileDetails [accessed 2024-12-20]
13. Schaeffer D, Berens EM, Vogt D. Health literacy in the German population. Dtsch Arztebl Int 2017 Jan 27;114(4):53-60 [FREE Full text] [doi: 10.3238/arztebl.2017.0053] [Medline: 28211318]
14. Schaeffer D, Berens EM, Vogt D, Gille S, Griese L, Klinger J, et al. Health literacy in Germany - findings of a representative follow-up survey. Dtsch Arztebl Int 2021 Oct 29;118(43):723-728 [FREE Full text] [doi: 10.3238/arztebl.m2021.0310] [Medline: 34551856]

15. Hurrelmann K, Klinger J, Schaeffer D. Gesundheitskompetenz der Bevölkerung in Deutschland: Vergleich der Erhebungen 2014 und 2020. Universität Bielefeld, Interdisziplinäres Zentrum für Gesundheitskompetenzforschung. 2020. URL: https://pub.uni-bielefeld.de/record/2950303 [accessed 2024-12-20]
16. Jordan S, Hoebel J. [Health literacy of adults in Germany: findings from the German Health Update (GEDA) study]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2015 Sep 31;58(9):942-950. [doi: 10.1007/s00103-015-2200-z] [Medline: 26227894]
17. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015 Dec 29;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
18. Wollmann K, van der Keylen P, Tomandl J, Meerpohl JJ, Sofroniou M, Maun A, et al. The information needs of internet users and their requirements for online health information-a scoping review of qualitative and quantitative studies. Patient Educ Couns 2021 Aug;104(8):1904-1932. [doi: 10.1016/j.pec.2021.01.020] [Medline: 33563502]
19. Röttele N, Schlett C, Körner M, Farin-Glattacker E, Schöpf-Lazzarino AC, Voigt-Radloff S, et al. Variance components of ratings of physician-patient communication: a generalizability theory analysis. PLoS One 2021;16(6):e0252968 [FREE Full text] [doi: 10.1371/journal.pone.0252968] [Medline: 34111197]
20. Radloff S, Wollmann K, Voigt-Radloff S, Farin-Glattacker E, Schlett C. Online-interventionen zur förderung der informiertheit. Impu!se für Gesundheitsförderung 2021;113(4):18.
21. van der Keylen P, Tomandl J, Wollmann K, Möhler R, Sofroniou M, Maun A, et al. The online health information needs of family physicians: systematic review of qualitative and quantitative studies. J Med Internet Res 2020 Dec 30;22(12):e18816 [FREE Full text] [doi: 10.2196/18816] [Medline: 33377874]
22. Voigt-Radloff S, Schöpf AC, Boeker M, Frank L, Farin E, Kaier K, et al. Well informed physician-patient communication in consultations on back pain - study protocol of the cluster randomized GAP trial. BMC Fam Pract 2019 Feb 25;20(1):33 [FREE Full text] [doi: 10.1186/s12875-019-0925-8] [Medline: 30803433]
23. van der Keylen P, Maun A, Hoyer S, Roos M, Beier M, Voigt-Radloff S, et al. "Mind the gap" — verzahnung von wissenschaft und hausärztlicher Praxis. Z Allg Med 2019;95:179-184 [FREE Full text] [doi: 10.3238/zfa.2019.0179-0184]
24. Braun C, Voigt-Radloff S. [Treatment of non-specific low back pain: evidence map of systematic reviews from 2015 to 2019]. Z Evid Fortbild Qual Gesundhwes 2019 Dec;149:12-21. [doi: 10.1016/j.zefq.2020.01.004] [Medline: 32151527]
25. Lang B, Kunzweiler K, Voigt-Radloff S. Gut informierte kommunikation zwischen arzt und patient im digitalen zeitalter – Das GAP-Projekt. In: Knieps F, Pfaff H, editors. Digitale Arbeit - Digitale Gesundheit. Berlin, Germany: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2017.
26. Schlett C, Röttele N, van der Keylen P, Schöpf-Lazzarino AC, Klimmek M, Körner M, et al. The acceptance, usability, and utility of a web portal for back pain as recommended by primary care physicians: qualitative interview study with patients. JMIR Form Res 2022 Dec 29;6(12):e38748 [FREE Full text] [doi: 10.2196/38748] [Medline: 36580365]
27. Nater T, Boyer C. Debate about evaluation and monitoring of sites carrying the HON-Logo. J Med Internet Res 2000 Jun 30;2(2):E13 [FREE Full text] [doi: 10.2196/jmir.2.2.e13] [Medline: 11720932]
28. Scheible R, Kadioglu D, Ehl S, Blum M, Boeker M, Folz M, et al. Enabling external inquiries to an existing patient registry by using the open source registry system for rare diseases: demonstration of the system using the European Society for Immunodeficiencies Registry. JMIR Med Inform 2020 Oct 07;8(10):e17420 [FREE Full text] [doi: 10.2196/17420] [Medline: 33026355]
29. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010;17(2):124-130 [FREE Full text] [doi: 10.1136/jamia.2009.000893] [Medline: 20190053]
30. Scheible R, Thomczyk F, Blum M, Rautenberg M, Prunotto A, Yazijy S, et al. Integrating row level security in i2b2: segregation of medical records into data marts without data replication and synchronization. JAMIA Open 2023 Oct;6(3):ooad068 [FREE Full text] [doi: 10.1093/jamiaopen/ooad068] [Medline: 37583654]
31. Scheible R, Strecker P, Yazijy S, Thomczyk F, Talpa R, Puhl A, et al. A multilingual browser platform for medical subject headings. Stud Health Technol Inform 2022 Jan 14;289:384-387. [doi: 10.3233/SHTI210939] [Medline: 35062172]
32. Strecker P, Boeker M, Buechner S, Scheible R. Usability evaluation of a modern multilingual MeSH browser. Stud Health Technol Inform 2022 Jun 29;295:37-40. [doi: 10.3233/SHTI220653] [Medline: 35773799]
33. Zowalla R, Wetter T, Pfeifer D. Crawling the German Health Web: exploratory study and graph analysis. J Med Internet Res 2020 Jul 24;22(7):e17853 [FREE Full text] [doi: 10.2196/17853] [Medline: 32706701]
34. Zowalla R, Pfeifer D, Wetter T. Readability and topics of the German Health Web: exploratory study and text analysis. PLoS One 2023 Feb 10;18(2):e0281582 [FREE Full text] [doi: 10.1371/journal.pone.0281582] [Medline: 36763573]
35. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
36. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
37. Nielsen J. Usability Engineering. Burlington, MA: Morgan Kaufmann Publishers; 1994.

38. Baumel A, Yom-Tov E. Predicting user adherence to behavioral eHealth interventions in the real world: examining which aspects of intervention design matter most. Transl Behav Med 2018 Sep 08;8(5):793-798. [doi: 10.1093/tbm/ibx037] [Medline: 29471424]
39. Mateus DA, Silva CA, de Oliveira AF, Costa H, Freire AP. A systematic mapping of accessibility problems encountered on websites and mobile apps: a comparison between automated tests, manual inspections and user evaluations. J Interact Syst 2021 Nov 08;12(1):145-171. [doi: 10.5753/jis.2021.1778]
40. Petrie H, Bevan N. The evaluation of accessibility, usability and user experience. University of York. 2009. URL: https://pure.york.ac.uk/portal/en/publications/the-evaluation-of-accessibility-usability-and-user-experience [accessed 2024-12-20]
41. Belinda BI, Kayode AB, Solomon N, Bethy TA. Analysis of internal and external website usability factors. Commun Appl Electron 2021 Apr;7(35):9-18. [doi: 10.5120/cae2021652880]
42. Kritz M, Gschwandtner M, Stefanov V, Hanbury A, Samwald M. Utilization and perceived problems of online medical resources and search tools among different groups of European physicians. J Med Internet Res 2013 Jun 26;15(6):e122 [FREE Full text] [doi: 10.2196/jmir.2436] [Medline: 23803299]
43. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
44. Zhang Y. Beyond quality and accessibility: source selection in consumer health information searching. J Assoc Inf Sci Technol 2014 Jan 07;65(5):911-927. [doi: 10.1002/asi.23023]
45. Saad M, Zia A, Raza M, Kundi M, Haleem M. A comprehensive analysis of healthcare websites usability features, testing techniques and issues. IEEE Access 2022;10:97701-97718. [doi: 10.1109/access.2022.3193378]
46. Fess - enterprise search server. Fess. URL: https://fess.codelibs.org/index.html [accessed 2021-04-07]
47. mozilla / readability. GitHub. URL: https://github.com/mozilla/readability [accessed 2021-04-07]
48. Tetrault D. Elastic app search: now available as a self-managed download. Elastic. 2019 Jun 25. URL: https://www.elastic.co/blog/elastic-app-search-7-2-0-released [accessed 2021-04-07]
49. Elasticsearch homepage. Elastic. URL: https://www.elastic.co/de/enterprise-search [accessed 2023-07-17]
50. Hashemian Nik D, Kasáč Z, Goda Z, Semlitsch A, Schulz S. Building an experimental German user interface terminology linked to SNOMED CT. Stud Health Technol Inform 2019 Aug 21;264:153-157. [doi: 10.3233/SHTI190202] [Medline: 31437904]
51. tala-med Suche. tala-med Suche Evidenzbasierte Gesundheitsinformationen. URL: https://suche.tala-med.info/ [accessed 2024-12-23]
52. WisoPanel homepage. WisoPanel. URL: https://www.wisopanel.net/ [accessed 2023-12-21]
53. Umfragesoftware. Unipark. URL: https://www.unipark.com/umfragesoftware/ [accessed 2024-12-23]
54. Soellner R, Huber S, Reder M. The concept of eHealth literacy and its measurement. J Media Psychol 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
55. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation in Industry. Boca Raton, FL: CRC Press; 1996.
56. Quirmbach SM. User experience und usability in suchmaschinen. In: Suchmaschinen: User Experience, Usability und Nutzerzentrierte Website-Gestaltung. Berlin, Heidelberg: Springer; 2012:57-65.
57. Magin M, Steiner M, Heinbach D, Bosold S, Pieper A, Felka EM, et al. Suchmaschinen auf dem Prüfstand – eine Vergleichende Inhaltsanalyse der Qualität von Trefferlisten. M&K 2015;63(4):495-516. [doi: 10.5771/1615-634x-2015-4-495]
58. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
59. Matomo homepage. Matomo. URL: https://matomo.org/ [accessed 2023-10-26]
60. Ripp S, Falke S. Analyzing user behavior with Matomo in the online information system Grammis. In: Proceedings of the XVIII EURALEX International Congress Lexicography in Global Contexts. 2018 Aug 29 Presented at: XVIII EURALEX; July 17-21, 2018; Ljubljana, Slovenia URL: https://ids-pub.bsz-bw.de/frontdoor/index/index/docId/7706
61. Shamat NA, Sulaiman S, Sinpang JS. A systematic literature review on user interface design for web applications. J Telecommun Electron Comput Eng 2017;9(3-4):57-61.
62. Mazumder FK, Das UK. Usability guidelines for usable user interface. Int J Res Eng Technol 2014 Sep;03(09):79-82. [doi: 10.15623/ijret.2014.0309011]
63. Nichols DM, McKay D, Twidale MB. Participatory usability: supporting proactive users. In: Proceedings of the 4th Annual Conference of the ACM Special Interest Group on Computer-Human Interaction. 2003 Presented at: CHINZ '03; July 3-4, 2003; Dunedin, New Zealand. [doi: 10.1145/2331829.2331841]
64. Davranova MD. Internet advertising: perceptions of the users. Int J Mark Bus Commun 2019;8(2 & 3):25-36 [FREE Full text]
65. Bildung in Deutschland (Bildungsstand). Destatis Statistisches Bundesamt. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/_inhalt.html [accessed 2023-12-08]
66. Kim S, Park K, Jo HS. Gap between perceived eHealth literacy and ability to use online cancer-related information. J Korean Med Sci 2020 Jun 22;35(24):e187 [FREE Full text] [doi: 10.3346/jkms.2020.35.e187] [Medline: 32567257]

67. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019 Presented at: NAACL-HLT 2019; June 2-7, 2019; Minneapolis, MN. [doi: 10.18653/v1/N19-1423]
68. Bressem KK, Papaioannou JM, Grundmann P, Borchert F, Adams LC, Liu L, et al. medBERT.de: a comprehensive German BERT model for the medical domain. Expert Syst Appl 2024 Mar;237:121598. [doi: 10.1016/j.eswa.2023.121598]
69. Lentzen M, Madan S, Lage-Rupprecht V, Kühnel L, Fluck J, Jacobs M, et al. Critical assessment of transformer-based AI models for German clinical notes. JAMIA Open 2022 Dec;5(4):ooac087 [FREE Full text] [doi: 10.1093/jamiaopen/ooac087] [Medline: 36380848]
70. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: a robustly optimized BERT pretraining approach. arXiv. Preprint posted online on July 26, 2019 [FREE Full text]
71. Scheible R, Frei J, Thomczyk F, He H, Tippmann P, Knaus J, et al. GottBERT: a pure German language model. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. 2024 Presented at: EMNLP; November 12, 2024; Miami, FL URL: https://aclanthology.org/2024.emnlp-main.1183 [doi: 10.18653/v1/2024.emnlp-main.1183]
72. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. Transact Assoc Comput Linguist 2017 Jun 1;5:135-146. [doi: 10.1162/tacl_a_00051]

Abbreviations

afgis: Aktionsforum Gesundheitsinformationssystem eV (Action Forum Health Information System eV)
BERT: Bidirectional Encoder Representations from Transformers
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
GAP: Gut informierte Kommunikation zwischen Arzt und Patient (Well-informed communication between physician and patient)
GDPR: General Data Protection Regulation
G-eHEALS: eHealth Literacy Scale, German version
i2b2: Integrating Biology and the Bedside
ISCED: International Standard Classification of Education
MeSH: Medical Subject Headings
RoBERTa: Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach
SERP: search engine results page
SNOMED-CT: Systematized Nomenclature of Medicine–Clinical Terms

Edited by E Borycki, Y Senathirajah; submitted 31.01.24; peer-reviewed by F Jahn, S Markham, L Hasan; comments to author 02.06.24; revised version received 12.07.24; accepted 23.10.24; published 23.01.25.

Please cite as:
Specht L, Scheible R, Boeker M, Farin-Glattacker E, Kampel N, Schmölz M, Schöpf-Lazzarino A, Schulz S, Schlett C, Thomczyk F, Voigt-Radloff S, Wegner C, Wollmann K, Maun A
Evaluating the Acceptance and Usability of an Independent, Noncommercial Search Engine for Medical Information: Cross-Sectional Questionnaire Study and User Behavior Tracking Analysis
JMIR Hum Factors 2025;12:e56941
URL: https://humanfactors.jmir.org/2025/1/e56941
doi:10.2196/56941
PMID:

Original Paper

Prioritizing Trust in Podiatrists' Preference for AI in Supportive Roles Over Diagnostic Roles in Health Care: Qualitative Interview and Focus Group Study

Mohammed A Tahtali[1,2*], MSc; Chris C P Snijders[2*], Prof Dr; Corné W G M Dirne[1*], Prof Dr; Pascale M Le Blanc[3*], Prof Dr

[1]Department of Industrial Engineering & Management, Fontys University of Applied Sciences, Eindhoven, The Netherlands

[2]Department of Industrial Engineering & Innovation Sciences, Human Technology Interaction group, Eindhoven University of Technology, Eindhoven, The Netherlands

[3]Department of Industrial Engineering & Innovation Sciences, Human Performance Management group, Eindhoven University of Technology, Eindhoven, The Netherlands

[*]all authors contributed equally

Corresponding Author:
Mohammed A Tahtali, MSc
Department of Industrial Engineering & Management
Fontys University of Applied Sciences
De Rondom 1, 5612 AP
Eindhoven
The Netherlands
Phone: 31 8850 79388
Email: m.tahtali@fontys.nl

Abstract

Background: As artificial intelligence (AI) evolves, its roles have expanded from helping out with routine tasks to making complex decisions, once the exclusive domain of human experts. This shift is pronounced in health care, where AI aids in tasks ranging from image recognition in radiology to personalized treatment plans, demonstrating the potential to, at times, surpass human accuracy and efficiency. Despite AI's accuracy in some critical tasks, the adoption of AI in health care is a challenge, in part because of skepticism about being able to rely on AI decisions.

Objective: This study aimed to identify and delve into more effective and acceptable ways of integrating AI into a broader spectrum of health care tasks.

Methods: We included 2 qualitative phases to explore podiatrists' views on AI in health care. Initially, we interviewed 9 podiatrists (7 women and 2 men) with a mean age of 41 (SD 12) years and aimed to capture their sentiments regarding the use and role of AI in their work. Subsequently, a focus group with 5 podiatrists (4 women and 1 man) with a mean age of 54 (SD 10) years delved into AI's supportive and diagnostic roles on the basis of the interviews. All interviews were recorded, transcribed verbatim, and analyzed using Atlas.ti and QDA-Miner, using both thematic analysis for broad patterns and framework analysis for structured insights per established guidelines.

Results: Our research unveiled 9 themes and 3 subthemes, clarifying podiatrists' nuanced views on AI in health care. Key overlapping insights in the 2 phases included a preference for using AI in supportive roles, such as triage, because of its efficiency and process optimization capabilities. There is a discernible hesitancy toward leveraging AI for diagnostic purposes, driven by concerns regarding its accuracy and the essential nature of human expertise. The need for transparency and explainability in AI systems emerged as a critical factor for fostering trust in both phases.

Conclusions: The findings highlight a complex view from podiatrists on AI, showing openness to its application in supportive roles while exercising caution with diagnostic use. This result is consistent with a careful introduction of AI into health care in roles, such as triage, in which there is initial trust, as opposed to roles that ask the AI for a complete diagnosis. Such strategic adoption can mitigate initial resistance, gradually building the confidence to explore AI's capabilities in more nuanced tasks, including diagnostics, where skepticism is currently more pronounced. Adopting AI stepwise could thus enhance trust and acceptance across a broader range of health care tasks, aligning technology integration with professional comfort and patient care standards.

(JMIR Hum Factors 2025;12:e59010) doi:10.2196/59010

KEYWORDS

AI's role in health care; decision-making; diabetes and podiatrists; trust; AI; artificial intelligence; qualitative; foot; podiatry; professional; experience; attitude; opinion; perception; acceptance; adoption; thematic; focus group

Introduction

Background

As artificial intelligence (AI) continues to advance, its potential to revolutionize health care is unquestionable. AI systems are increasingly capable of performing tasks that range from routine data analysis to complex decision-making processes that were once the sole domain of human experts. These technologies offer the promise of more accurate predictions [1], more precise patient condition explanations [2], autonomous diagnoses [3], and personalized treatment plans [4]. Despite these capabilities, there remains an evident hesitancy among health care professionals and patients alike to embrace AI-driven recommendations fully. This skepticism is partly rooted in worries about AI's capacity to grasp patients' nuanced, individual scenarios [5], to integrate the crucial human elements of empathy and understanding [6,7], and to deal with complex ethical considerations in health care [8].

The literature on trust in AI highlights concerted efforts to foster trust by exploring strategies to address trust issues between AI systems and users. For example, some studies suggest that redefining AI as a supportive adjunct to professionals rather than a standalone solution could promote acceptance [9,10]. In addition, showcasing instances where AI surpasses human experts in specific tasks may enhance humans' trust in AI [5,9,11]. Emphasizing transparency and providing understandable explanations of AI processes are crucial, with evidence showing users how AI works can greatly diminish skepticism [12-15]. For example, applying explainable AI (XAI) techniques to explain the basis of AI-generated predictions can, in certain contexts, enhance human trust and understanding [12]. Although XAI can potentially be valuable in increasing trust and understanding, the specific implementation of XAI (eg, the complexity and relevance of the explanation) is crucial in achieving those goals [16]. However, despite these positive developments, fully integrating AI into practice remains challenging, and professionals and laypersons often still prefer human judgment and intuition over AI recommendations [6,7,10,17].

AI's potential applications within health care are vast and multifaceted. Its well-established ability to detect diseases such as cancer and predict various health conditions highlights its substantial role in diagnostics. Not only does AI excel in recognizing patterns within medical images for early disease detection, offering a level of precision that can sometimes surpass traditional methods [3,18,19], but also its influence extends beyond diagnostics. AI assumes a crucial supportive role in health care operations, such as triaging in emergency departments where algorithms prioritize care based on urgency and outcomes, thereby optimizing resource allocation and ensuring timely patient attention [20]. Furthermore, AI's capacity for enhancing operational efficiency is evident across a spectrum of health care activities. It predicts bed availability and automates the conversion of patient consultations into concise summaries, thereby freeing health care professionals from administrative work. This enables more patient-centered care and marks AI as a transformative tool in health care [21].

This study seeks to delve deeper into health care providers' perspectives on AI's current and potential role and value in their field. While previous studies often focus on assessing trust in already developed AI systems, this study takes a different approach by involving health care professionals from the outset in identifying potential applications of AI. Early engagement allows us to determine where AI is anticipated to be most valuable, fostering a more collaborative and user-centered integration that aligns with professionals' needs and expectations. By engaging directly with health care professionals who are the prospective users of these AI applications, the study seeks to pinpoint areas where AI integration would be most advantageous and eagerly anticipated by practitioners. Through this inquiry, we aspire to pave the way for more effective and user-endorsed integration of AI in health care, aligning technological advancements with health care professionals' real-world needs and preferences.

Objective

Our study explores health care providers' perceptions of AI's potential benefits in patient care and operational efficiency, with examples including risk assessments and diagnostics. The effectiveness of AI technologies depends on their advanced capabilities and, more crucially, on their acceptance by health care professionals. Without their acceptance, achieving a truly successful implementation becomes only partial, leading to the largely unrealized potential of AI. Thus, we aim to understand these professionals' perspectives on AI adoption, emphasizing the development of trust and integration strategies and identifying approaches that position AI as a valuable tool complementing human expertise, not replacing it [10]. Consequently, the central research question we propose is as follows: How can AI be effectively integrated into practice in a way that ensures health care professionals' acceptance and trust in its advice?

We adopt a qualitative approach to address our research question, focusing on the FootCheck (RondOm Podiatrists) app (in Dutch: Voetencheck app). This app, specifically designed for the daily monitoring of patients with diabetes in the Netherlands, sends daily push notifications to guide these patients through a foot screening process, featuring instructional videos and questions. This enables patients to submit photos to podiatrists to evaluate potential issues such as wounds or ulcers. During its successful pilot phase, which occurred amid the COVID-19 pandemic, the app was used by over 1700 individuals, leading to more than 43,000 screenings [22]. Recognizing the growing interest of stakeholders in leveraging

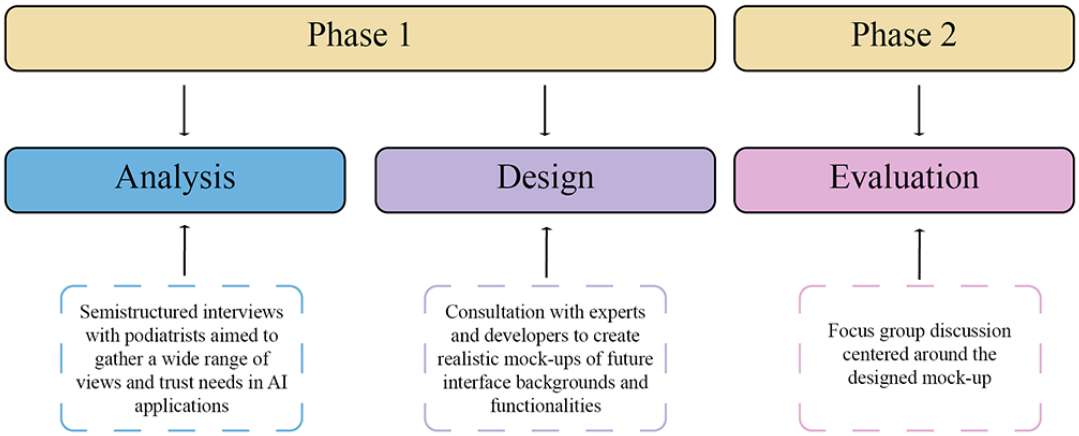
AI to enhance efficiency and health care quality, we believe the FootCheck app provides an ideal basis for exploring AI's potential. It is important to note that the app still lacks AI integration, highlighting the potential benefits of AI for future enhancements. Our study includes interviews with diverse podiatrists, encompassing current and potential future app users. These interviews aim to understand their experiences and perspectives regarding using the app as it is and the prospects for integrating AI into the screening process of future versions of the app.

Methods

Design and Setting

We conducted a qualitative study structured in 2 phases to investigate trust and acceptance in AI. In phase 1, we sought to capture diverse perspectives and needs from podiatrists regarding their trust in and acceptance of AI technologies through semistructured interviews. The insights gathered from these interviews served as input to the mock-up design process. Phase 2 aimed to further explore specific AI applications through a focus group session. Figure 1 provides a visual representation of these phases.

Figure 1. A visual presentation of the 2 phases of the study. AI: artificial intelligence.



The FootCheck App and Decision-Making Process

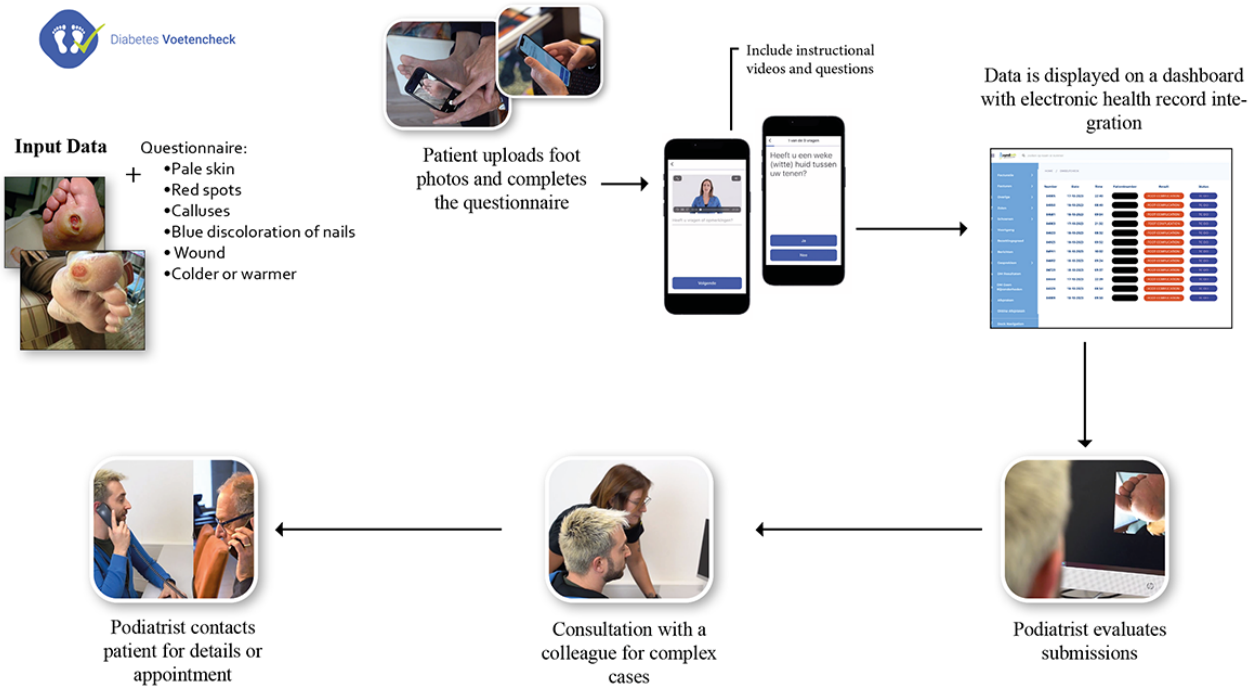
The practical application of AI in health care settings we studied was the FootCheck app, developed for RondOm Podiatrists. As one of the largest podiatry chains in the Netherlands, RondOm Podiatrists introduced this app during the COVID-19 pandemic to ensure continuous care for their patients. The app has since been used to facilitate daily monitoring of foot health for patients with diabetes, allowing patients to submit photographs for screening. The workflow from patient submission to professional evaluation is illustrated in Figure 2.

RondOm Podiatrists are keenly interested in the potential of AI to enhance their app and streamline decision-making processes. The system does not use AI currently, but there are numerous potential areas where AI could provide substantial benefits. For example, AI could be instrumental in assessing the suitability of submitted photos for evaluation. This capability would significantly reduce the workload on professionals by automatically filtering out low-quality or irrelevant images, ensuring that only those suitable for review are presented. This efficiency gain is crucial as it allows podiatrists to focus their expertise where it is most needed rather than on preliminary sorting tasks.

Moreover, AI could assist in prioritizing cases on the basis of urgency. Currently, podiatrists must manually open and evaluate all submitted photos, which is time-consuming and potentially less effective in quickly identifying urgent cases. By implementing AI to perform a preselection of instances based on predetermined criteria of urgency, the system could ensure that patients requiring immediate attention are identified and prioritized, thereby optimizing patient care and resource allocation.

Another critical area in which AI could have a significant role is aiding in the diagnosis of foot conditions. The increasing number of patients with diabetes and a relative shortage of podiatrists present a growing challenge. AI can alleviate some of this burden by providing preliminary assessments based on the analysis of photos. While the professional community may have reservations about relying entirely on AI for diagnoses, its role in offering support and preliminary assessments could be invaluable. This assistance could help bridge the gap in service provision, ensuring that more patients receive timely and effective care.

Figure 2. Workflow from patient submission to professional evaluation.



Phase 1: Participants and Recruitment Process

In the first phase of our study, we engaged in interviews with 9 podiatrists from 4 distinct practices, aiming to gather diverse insights into the use and perception of the health care app in question. The actual use of the current app varied considerably among participants, encompassing some who used the app daily and others who, while not yet current users, were familiar with its functionalities. Furthermore, participants' professional expertise spanned different aspects of podiatric care, including both foot care for patients with diabetes and general foot health services.

The recruitment of participants was facilitated through a collaborative effort with the Fontys University of Applied Sciences, one of the leading institutions for podiatry education in the Netherlands. This collaboration was essential for 2 reasons. First, Fontys University maintains strong ties with its alumni, many of whom stay actively engaged with the university through professional networks, alumni events, and workshops. Second, Fontys University's Health Innovations & Technology research group maintains regular collaborations with health care institutions, ensuring strong connections within the field. Through these partnerships, we recruited a diverse group of motivated podiatrists who voluntarily participated in the study. By using Fontys University's network and its respected position within the field, we ensured the participation of podiatrists with varying levels of experience and expertise across different areas of podiatric care, contributing to a representative sample for our study. To accommodate our participants' varying schedules and preferences, the interviews were conducted using a hybrid approach; some were held in person, taking advantage of the natural setting of the participants' practice environments, while others were carried out through the Teams (Microsoft Corp) platform.

Data Collection

We started an extensive literature review on trust in AI, both generally and within the health care context, from which we identified key topics. These topics were categorized into 4 groups as follows: knowledge of AI, trust in humans and AI advice, AI's added value to daily tasks, and factors affecting trust in AI. On the basis of these categories and codes, we formulated initial interview questions. Next, we sought feedback on these initial questions from experts in human and machine interaction. We incorporated their feedback and refined the questions to ensure their relevance and comprehensiveness. To validate the clarity and comprehensiveness of these questions, we conducted a preliminary interview with an experienced podiatrist. This pilot interview served to ensure the questions were understandable and relevant for the prospective interviewees, and it also provided an opportunity to identify any missing questions from the practical perspective of podiatry. The resulting interview guide is detailed in Multimedia Appendix 1.

Participants were briefed on the study's objectives—namely, to explore their perceptions of AI, identify factors influencing trust in AI among podiatrists, examine the impact of trust on AI adoption, and assess their needs for trusting AI in patient care. They were requested to share demographic information via a link before the interview sessions. The specific interview questions were withheld in advance to preserve the authenticity of responses. The interviews were conducted from July to September 2023.

Consent for recording was obtained from all podiatrists, with the sessions captured either on a voice recorder or through the Teams platform, depending on the interview format. These recordings were transcribed verbatim, using Amberscript [23] for efficiency and accuracy. For interviews conducted on the

Teams platform, we leveraged the platform's transcription functionality. Each interview spanned approximately 30 to 45 minutes, with podiatrists having provided explicit consent, either written or digital, for participating and recording their responses. They were also assured of their right to withdraw from the study at any time.

Data Analysis of the Semistructured Interviews

We used a dual-coding strategy, where 2 researchers independently analyzed the interview transcripts to ensure a robust examination of the data. Researcher 1 (MAT), a PhD candidate with a solid background in qualitative research methodologies, and researcher 2 (Ruben Gloudemans) used distinct software tools for coding—Atlas.ti (version 11; Atlas.ti Scientific Software Development GmbH) and QDA-Miner (version 6.0; Provalis Research). This divergence in tool use stemmed from their prior experiences and the software licenses available.

Following the thematic analysis by Braun and Clarke [24], we initiated our analysis by thoroughly reading the transcripts to gain an initial understanding and verify their accuracy. First, we labeled the transcripts without predefined categories, using an inductive approach leading to an extensive list of initial codes directly derived from the data. For example, codes such as human interaction, urgency ranking, efficiency, risk detection, and false advice were formed. We then examined these initial codes to determine if they overlapped with the topics we had previously identified from the literature and combined them if needed. For example, urgency ranking and risk detection were put together since both are related to the topic added value of AI. In this step, we started a deductive approach, where we applied existing theoretical frameworks to refine our codes into coherent themes. We repeated this analytical process several times, refining our approach each time to ensure the emergence of concrete themes. This process yielded a preliminary set of themes.

Subsequently, our individually identified themes were compared in a meeting, discovering significant overlap. For some themes where our interpretations initially diverged, we reached consensus through discussion, using inductive and deductive reasoning. These discussions revealed that, despite using different terminologies, we were referring to the same thematic content. Our thematic analysis was refined further through this iterative process of inductive and deductive coding.

Mock-Up Design

Guided by our research question on how AI can effectively integrate into the FootCheck app's existing screening process with user acceptance, we sought the podiatrists' insights on how they would like to receive advice and in which forms to ensure their confidence in it. Various suggestions emerged, such as explaining the reasoning behind the advice, photos pinpointing potential complications on the feet, and color-coding systems for triage. However, the first phase did not conclusively determine which forms of applications and the level of transparency would foster greater trust.

To further explore these possibilities, we organized a focus group featuring mock-ups. Presenting a realistic representation was crucial, particularly considering that currently, the app and its screening process do not use AI.

1. The first mock-up was designed to assess the effectiveness of AI in a supportive role, specifically in scheduling and prioritizing patient consultations based on an array of time indicators.
2. The second mock-up aimed to explore the potential of AI in diagnosing conditions by analyzing photographs submitted by users and identifying common foot complications such as ulcers or calluses.

We aimed to delve deeper into assessing these AI roles by examining their functionality and acceptability among podiatrists. We crafted 3 distinct design variations for each mock-up to achieve this, guided by needs and concerns expressed in the first interviews with the podiatrists. Our primary reason for this approach was to facilitate a broad discussion by presenting transparency levels varying from low to medium to high. This method allowed us to not only compare the discussions across different AI applications but also delve into the discussions within each level of transparency. Thereby, we aimed to explore the range in podiatrists' perceptions and acceptance extensively.

In designing these mock-ups, we used the current interface environment to ensure familiarity for users. In addition, we consulted with experts during the design process to ensure the mock-ups were practical and reflective of current best practices in the field. For an illustration of mock-up 1 with a high transparency level, see Figure 3; for the diagnostic mock-up set in a medium transparency, refer to Figure 4. For a detailed overview of all variants, refer to Multimedia Appendix 2.

Figure 3. Mock-up 1: high-transparency variant.

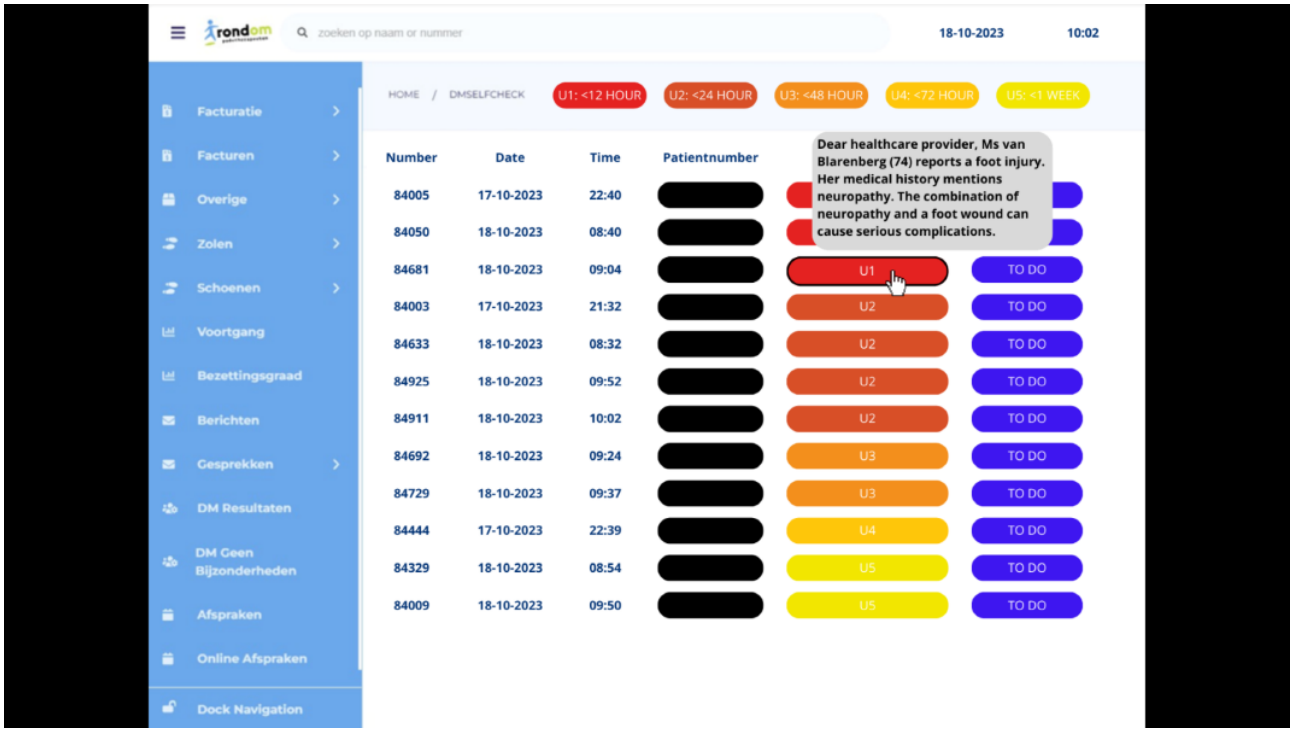
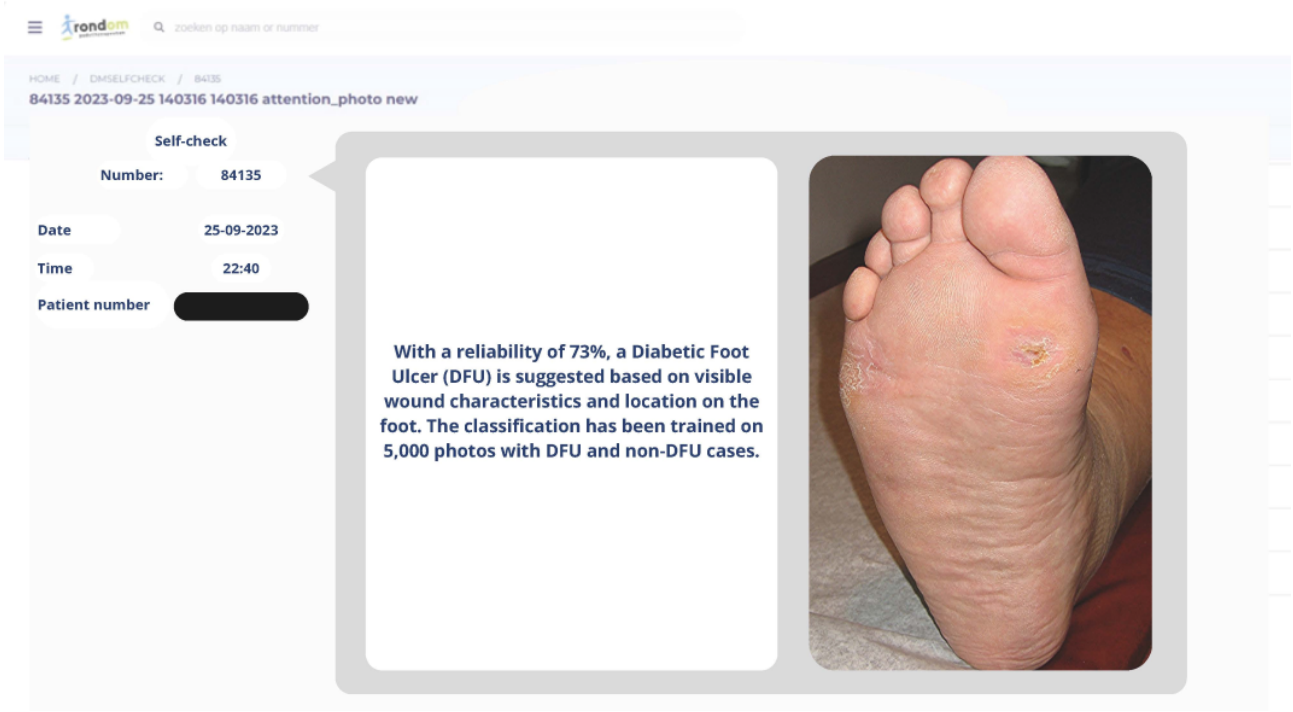


Figure 4. Mock-up 2: medium-transparency variant.



Phase 2: Participants and Recruitment Process

Phase 2 was aimed at enriching our understanding and validating the outcomes of the first phase. Initially, we used the same approaches that had effectively recruited participants for phase 1, including relying on Fontys University's network and existing partnerships within the podiatry field. We also reached out to contacts from our previous interviews. However, these efforts alone were insufficient to meet our needs for new participants for phase 2. To broaden our reach, we engaged with potential candidates on social media platforms such as LinkedIn (LinkedIn Corp) and Facebook (Meta Platforms, Inc) and attended podiatrist conferences to connect with new professionals. Through these combined efforts, we recruited 4 new podiatrists who had not been involved in the first phase and reinvited one participant from phase 1. The focus group session was organized at the Fontys University of Applied Sciences' inspiration space in December 2023.

Procedure for Focus Group and Data Collection

To evaluate our mock-ups, we created an interview guide on essential topics derived from the findings of phase 1: trust in AI's role, transparency preferences, and AI's effect on decision-making, outlined in Multimedia Appendix 3. During the focus group session, participants were thoroughly briefed on the structure of the session that was divided into two 45-minute segments with a break in between. The first session explored mock-up 1, while the second delved into mock-up 2.

The roles of MAT and Ruben Gloudemans within the session were clearly defined for participants; MAT served as the observer, summarizing discussions and posing questions, whereas Ruben Gloudemans led the session and presented the materials. Participants provided written consent by signing a consent form.

After a brief introduction, participants were provided two A3-sized sheets, each showcasing 3 mock-up variants highlighting different transparency levels. Before initiating the group discussions, participants were asked to individually rank these variants from 1 to 3, with 1 being their top preference, and to write down comments next to each variant explaining their choices. This step was designed to minimize mutual influence and ensure that each participant's initial reactions and thoughts were captured independently. Following this individual reflection, detailed group discussions were held, focusing on participants' rationale regarding their preferences and trust in the AI advice shown in each variant.

Data Analysis Focus Group

All participants attended in person for the data collection of phase 2. Initially planned as two 45-minute segments with a 15-minute break in between, the session ultimately lasted 70 minutes in a single stretch. This was shorter than anticipated, primarily because the first segment was concluded earlier than expected, and participants opted to proceed without taking the scheduled break.

We adopted a more structured coding method as we progressed to the second phase, building on the codes and themes identified in the first phase. This shift was enabled by implementing a thematic framework approach, a systematic strategy that helped us to systematically organize and interpret the data [25].

After verifying transcription accuracy and familiarizing ourselves with the transcript, we developed an analytical framework. This framework included predefined themes from phase 1: AI's roles, transparency levels, data reliability, trust development, and the significance of human interaction. This step laid the groundwork for indexing, where we systematically tagged the data with our framework's themes. Next, through charting, we distilled participants' responses into structured summaries, aligning them with the thematic framework. In the indexing and charting phase, we organized the data into themes linked to our research goals, creating visual charts. This approach helped us spot patterns and differences, uncovering insights that enhanced our understanding.

Ethical Considerations

No medical devices were used in the studies, and no physical interventions were performed on participants. The image used in mock-up 2 originated from publicly available internet sources and was appropriately licensed. Consequently, under regulations stipulated by the Dutch Medical Research Involving Human Subjects Act (WMO), the necessity for ethics approval from a national ethics committee was circumvented. However, ethical clearance was proactively obtained from the internal ethics committee of Fontys University of Applied Sciences, documented under project numbers 050324 and 080324. There were no conflicts of interest between the participants and the FootCheck app. None of the participants had any financial, developmental, or other professional ties to the app.

Results

Phase 1: Semistructured Interviews

Participants

Participant demographics for phase 1 are listed in Table 1. In total, 9 individuals, all with backgrounds in podiatry, participated in the first phase of our study. A distinguishing characteristic of this cohort is their active engagement in clinical practice, that is, their extensive practical experience.

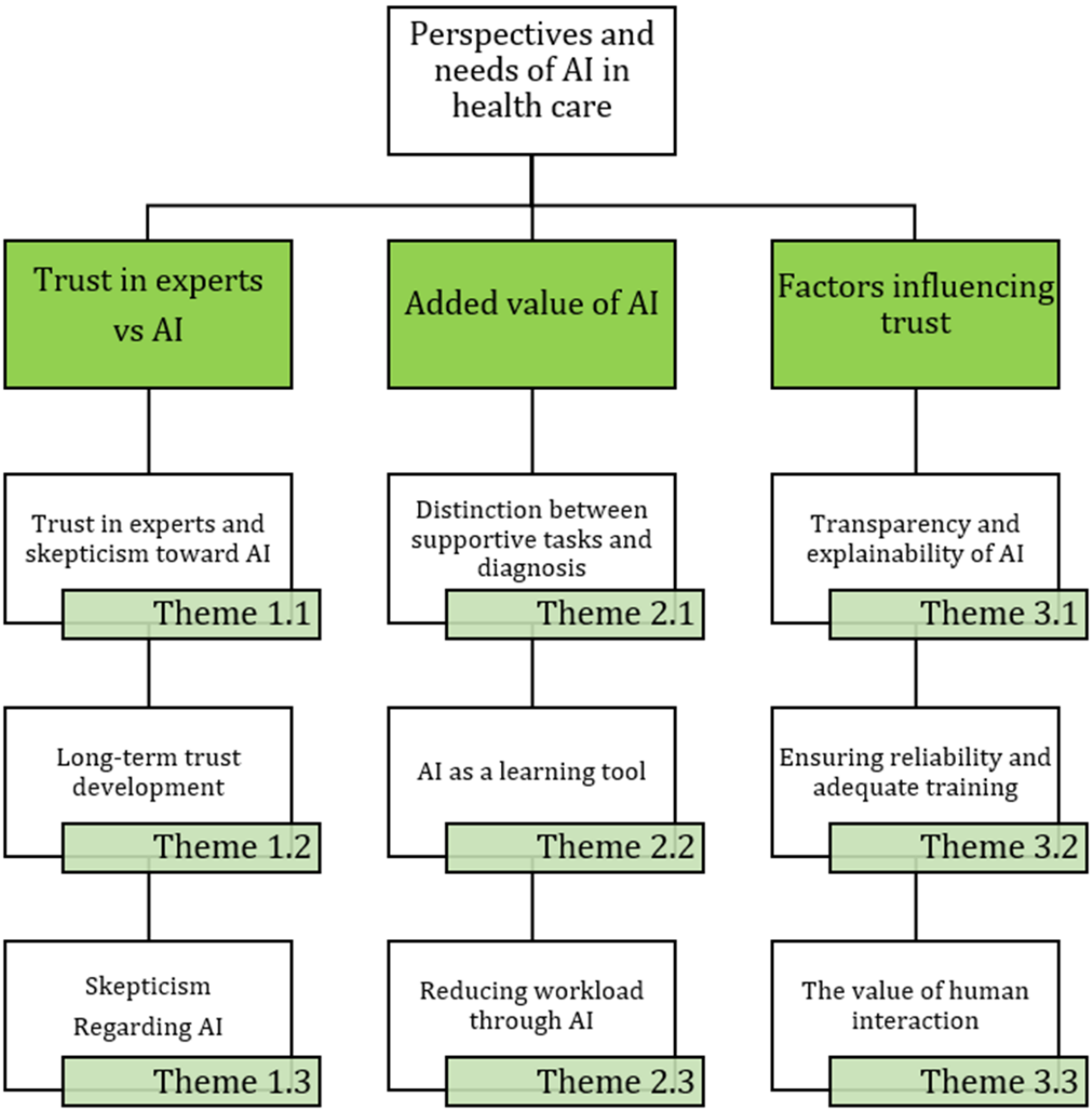
Table 1. Participants' characteristics from semistructured interviews in phase 1 (n=9).

Characteristics	Participants, n (%)
Sex	
Female	7 (78)
Male	2 (22)
Age group (y)	
20-30	1 (11)
31-40	2 (22)
41-50	3 (33)
51-60	2 (22)
61-70	1 (11)
Education level	
Bachelor's	4 (44)
Master's	4 (44)
PhD	1 (11)
Specialized education: foot care for patients with diabetes	
Yes	4 (44)
No	5 (56)
Experience (mo)	
0-6	1 (11)
6-12	2 (22)
12-24	4 (44)
24-36	2 (22)
Experience with the app	
Yes	2 (22)
No	7 (78)

Themes From the Interviews

The interviews with podiatrists revealed 3 critical themes regarding podiatrists' perspectives and requirements for AI integration in the FootCheck app and related decision-making process: trust in human experts versus in AI, the added value of AI, and the factors influencing trust in AI. Each theme is further divided into 3 subthemes. A detailed overview of these key themes can be found in Figure 5.

Figure 5. Generated themes and subthemes from phase 1: perspectives of experienced podiatrists on the FootCheck app and decision-making process. AI: artificial intelligence.



Theme 1: Trust in Experts Versus AI

In the initial phase of the interview, we began by asking participants about their understanding of AI and their trust in it compared with their colleagues (ie, experts). We aimed to uncover how AI impacts their trust levels and the underlying reasons for these views.

Theme 1.1: Trust in Experts and Skepticism Toward AI

Podiatrists expressed trust in the advice from expert colleagues, on the basis of their extensive knowledge and experience. They recognized that making mistakes is part of being human and considered it normal for colleagues to err every now and then. Conversely, when receiving advice from AI, direct trust is not immediately established for most podiatrists. They maintained a critical stance, emphasizing the need to verify the accuracy of AI-generated advice:

> *I trust a colleague, sometimes blindly, because they have much experience.* [Participant 7, female, age group 61-70 years]
>
> *I see the added value of AI advice. But I still want to check it or have something alongside it.* [Participant 2, female, age group 51-60 years]
>
> *I do have trust in advice from colleagues. I've considered practical situations, and nine times out of ten, I agree with them just fine. And there are always nuanced differences because you see things slightly differently with your experience, perspective, and view. But let me put it this way: I firmly trust it.* [Participant 8, male, age group 41-50 years]

Theme 1.2: Long-Term Trust Development

Participants indicated that AI may initially struggle to provide accurate advice due to the limited data available and the learning phase required for the model. They also mentioned their initial lack of experience with using AI, which leads to uncertainty about its performance. However, participants felt their trust can increase as they gain more experience with the AI model:

> *In the beginning, I was sceptical, but as it sees and evaluates more photos, my trust grows.* [Participant 9, female, age group 20-30 years]
>
> *Yes, AI needs to prove itself first, and then it can add value.* [Participant 7, female, age group 61-70 years]
>
> *Indeed, my trust in AI would significantly increase if, over time, the technology evolves to a stage where the AI system is fully operational and applicable.* [Participant 6, female, age group 41-50 years]

Theme 1.3: Skepticism Regarding AI

Participants expressed significant concerns about the potential errors an AI system might make and the severe consequences that could arise from such mistakes. In addition, they highlighted a notable limitation of AI: its inability to conduct physical examinations on patients. Such examinations are often essential for a thorough understanding of a patient's foot health, indicating a gap in AI's capabilities in health care settings:

> *Yes, if a patient says: "I have pain in my foot, for example, in the ankles." Well, AI can't see that or physically feel or move it.* [Participant 9, female, age group 20-30 years]
>
> *We're not talking about a patient getting advice to buy Shoe A, and then they say it's too big and buy another shoe.... We're talking about serious outcomes, like the loss of a foot, a leg, or even a life, and that's a significant responsibility.* [Participant 4, female, age group 31-40 years]
>
> *The consequences are quite significant if you miss something there.* [Participant 7, female, age group 61-70 years]

Theme 2: Added Value of AI

Overview

In exploring AI's potential impact on podiatry, we engaged participants in discussions about integrating AI into their everyday practice. This dialogue revealed 3 insightful subthemes, each shedding light on AI's role in enhancing podiatric care. The following sections delve into these emergent themes.

Theme 2.1: Distinction Between Supportive Tasks and Diagnosis

Nearly all participants agreed that AI should be used for support tasks such as triage. They highlighted that significant time is spent retrieving missing information not provided by patients and addressing issues with incorrect photo submissions that impede accurate assessment. They were of the opinion that these tasks could be efficiently managed by AI, thereby streamlining the process and improving overall efficiency. However, opinions varied on whether AI can diagnose foot conditions in patients with diabetes. Most participants were of the opinion that AI is incapable to do so, emphasizing the need for physical examinations by human experts to come up with an accurate diagnosis. They pointed out that photos fail to reveal crucial details, such as hidden wounds beneath calluses, which are not directly visible in pictures. Furthermore, participants noted the importance of physically assessing the feet for swelling and temperature variations (ie, warmth or coldness) that cannot be discerned from photos alone:

> *If it can highlight specific areas of concern, suggesting "pay attention here, this is alarming," then I do place my trust in it.* [Participant 3, female, age group 41-50 years]
>
> *AI can't make complete diagnoses; you need to hold the foot in your hands.* [Participant 2, female, age group 41-50 years]
>
> *I'm considering integrating a feature for predicting situations where an immediate consultation is crucial and distinguishing those from cases where it might not be as urgent.* [Participant 1, male, age group 41-50 years]
>
> *If it integrates well with practical needs, like photos can be easily uploaded directly into the file, making follow-up simpler. Yes, so it actually assists in your work.* [Participant 6, female, age group 41-50 years]

Theme 2.2: AI as a Learning Tool

Participants acknowledged the potential of AI as a valuable learning tool, particularly for identifying unfamiliar diseases and suggesting innovative treatment strategies. They expressed optimism about AI's ability to broaden their diagnostic and therapeutic horizons:

> *Maybe the AI will uncover a disease pattern I haven't encountered before, offering me an opportunity to expand my knowledge.* [Participant 1, male, age group 41-50 years]
>
> *Hey, AI sees things differently, and that's something I can learn from. I do see the added value in that.* [Participant 5, female, age group 20-30 years]
>
> *In this light, I believe that when an AI indicates, "this is the optimal treatment," it becomes incredibly fascinating. It prompts a collective inquiry: "Why hadn't this occurred to us before? Is this indeed the correct approach? Or is there already a hospital implementing this method?"* [Participant 4, female, age group 31-40 years]

Theme 2.3: Reducing Workload Through AI

Participants highlighted the health care sector's challenge of increasing workloads, worsened by a decreasing number of podiatrists and a growing number of patients with diabetes. There was widespread agreement on AI's ability to alleviate this load. By automating processes such as triage, AI can significantly reduce the workload, freeing podiatrists to focus on the more crucial parts of patient care:

As the pressure continues to increase, I see AI as a great tool to indeed perform a kind of triage to ensure that the workload doesn't become overwhelming. [Participant 3, female, age group 41-50 years]

But yes, to be completely honest, looking towards the future, we definitely need to reduce the workload; otherwise, we'll end up with even fewer people having to help more individuals. [Participant 4, female, age group 31-40 years]

Theme 3: Factors Influencing Trust

Overview

Finally, we delved into the pivotal elements that shape trust in AI-provided advice. We also probed when participants might be inclined to defer to AI's judgment over their own. This investigation unveiled 3 subthemes, each shedding light on the interaction between humans and AI systems.

Theme 3.1: Transparency and Explainability of AI

All participants agreed that clear explanations and transparency are essential, considering them as fundamental to trusting AI. They wished to comprehend the methodology AI uses to arrive at its recommendations. They proposed several methods to enhance transparency, such as providing concise explanations of the criteria AI uses for its evaluations, statistical performance data, or visual aids such as photos with overlaying layers or color coding:

I believe factors like the colour of the callus, the size of the area, its location, and history also play a role. Has there been a wound in that spot before? Yes, things like that. [Participant 8, male, age group 41-50 years]

For example, I want to see a number of photos of the foot along with some explanation about them. I want to understand how the evaluation system is designed and the data it uses. [Participant 2, female, age group 51-60 years]

It increases my confidence in knowing the sources from which AI retrieves its information or the ones I've provided to it. [Participant 1, male, age group 41-50 years]

Theme 3.2: Ensuring Reliability and Adequate Training

Participants were of the opinion that AI should be advanced enough from the beginning, emphasizing that a more extensive database of photos would increase their trust in the system. They would like AI to provide consistent advice, ideally on par with a fellow podiatrist. Training was highlighted as a crucial element for enhancing AI's performance. In addition, understanding who the designers are and the data used for training is considered necessary. If participants know the designers are experts in their field, they feel more confident in the AI tool's reliability:

Well, if you've only input three things into it, I won't trust it. But if you've input 300,000 things, then at some point, you really know better. [Participant 2, male, age group 41-50 years]

Well, I'm not quite sure how that works, but if that's the case, then the people involved in setting it up need to have the right background knowledge, the right experience with the subject; they need to be specialists, so to speak. [Participant 2, female, age group 51-60 years]

Theme 3.3: The Value of Human Interaction

Participants emphasized that future AI applications must not compromise their interaction with patients, highlighting each patient's uniqueness and background. They cautioned that AI may not always account for these personal nuances, underlining the importance of maintaining the human touch in health care:

I still believe that personal human contact will never disappear. [Participant 3, female, age group 41-50 years]

All those kinds of things. Yes, they're slightly different for everyone. [Participant 1, male, age group 41-50 years]

But you still maintain human contact between the client and yourself, allowing them to easily ask additional questions, for instance, to you. [Participant 5, female, age group 61-70 years]

Moving From Phase 1 to Phase 2

The findings from phase 1 revealed that AI can significantly improve various practical aspects of podiatry, particularly by alleviating workloads and serving as a learning tool. A clear preference was observed for using AI in supportive tasks rather than diagnostic roles. This reflects a cautious attitude of podiatrists toward AI-generated advice and raised the question: Will AI support be more readily accepted in supportive than in diagnostic tasks, and what factors influence this difference in acceptance?

Therefore, phase 2 of our study sought to explore the specific applications of AI in supportive versus diagnostic roles within podiatry in detail. Given the ongoing reluctance toward its role in diagnostics, this study aimed to understand the elements that can foster trust in AI for supportive tasks. By introducing a new group of podiatrists, phase 2 of our study aimed to gather fresh insights and assess whether the findings observed in phase 1 align with new professional perspectives.

Phase 2: Focus Group

Participants

Table 2 outlines an overview of the demographics of the focus group participants. The group consisted of 5 participants, 4 of whom were new podiatrists who engaged in the discussion. This group consisted of specialists in foot care for patients with diabetes, who primarily provide daily consultations to these patients.

Table 2. Participants' characteristics from focus groups in phase 2 (n=5).

Characteristics	Participants, n (%)
Sex	
Female	4 (80)
Male	1 (20)
Age group (y)	
41-50	2 (40)
51-60	2 (40)
61-70	1 (20)
Education level	
Bachelor's	2 (40)
Master's	3 (60)
Specialized in foot care for patients with diabetes	
Yes	5 (100)
No	0 (0)
Experience (y)	
20-30	4 (80)
31-40	1 (20)
Experience with the app	
Yes	2 (40)
No	3 (60)

In phase 1 of our study, we opted for a flexible approach, allowing themes to emerge directly from the transcribed data. This exploratory phase enabled us to capture many insights directly from participants' experiences and perspectives. Moving into phase 2, we shifted toward a more structured approach, using themes, which were relevant for phase 2 and predefined in phase 1. This transition allowed us to refine our analysis, focusing on the detailed examination of specific, predetermined themes to gain deeper insights and understanding of the topics at hand.

For a clear overview, we first present the key findings from the participants about various transparency levels for the 2 mock-ups in Table 3. Subsequently, in Table 4, we display the other themes carried forward from phase 1 for further analysis in phase 2, ensuring a comprehensive and structured exploration of the data.

Table 3. Thematic framework analysis for key findings and preferences of transparency levels in AI[a] applications.

Theme: transparency and explainability of AI	Mock-up 1: AI support (triage and scheduling)	Mock-up 2: AI diagnosis (photo analysis)
Low (variant A)	Podiatrists perceive this variant as a "black box," with unclear determination of urgency and relevance of the presented information	Feedback indicates general dissatisfaction and a lack of clarity on how AI reaches its conclusions
First preference	—[b]	P4 and P5 (n=2)
Second preference	—	—
Third preference	P1-P5 (n=5)	P1-P3 (n=3)
Medium (variant B)	Podiatrists express concerns over the reliability of patient feedback as input for AI, pointing out a lack of crucial information such as wound location	While the text is deemed acceptable, this variant raises questions about the necessity of repeating information
First preference	—	—
Second preference	P1-P5 (n=5)	P1-P5 (n=5)
Third preference	—	—
High (variant C)	Preference is given to detailed information, including specific medical details and clear patient identification	Podiatrists appreciate the clarity and added layer of understanding this variant provides but question the necessity of certain statistics, such as a reliability percentage
First preference	P1-P5 (n=5)	P1-P3 (n=3)
Second preference	—	—
Third preference	—	P4 and P5 (n=2)

[a]AI: artificial intelligence.

[b]Not applicable.

Table 4. Thematic framework analysis for AI[a] applications and predefined themes.

Themes	Participant 1 (female; age group 51-60 y)	Participant 2 (female; age group 41-50 y)	Participant 3 (male; age group 41-50 y)	Participant 4 (female; age group 51-60 y)	Participant 5 (female; age group 61-70 y)
Distinction between supportive tasks and diagnosis	Prefers AI for triage due to speed and accuracy. Skeptical about diagnostic capability without physical examination	—[b]	Values AI's ability to prioritize photos based on urgency, enhancing patient care efficiency	Believes AI is incapable of handling complex diagnostic tasks, such as differentiating between types of wounds	Asserts that computers cannot replace human nuanced observation in diagnostics, emphasizing AI's use in triage
Ensuring reliability and adequate training	Values the volume of feedback, similar to how reviews inform vacation choices	Trust is reinforced by a large volume of photos, indicating data reliability	Importance of data diversity, including training AI on poor-quality photos	Questions the repetitive mention of the photo database size suggest it is only necessary in training	Trust depends on knowing the AI's advice is based on an extensive and diverse photo database
Long-term trust development	—	Trust in the algorithm grows over time with proven accuracy	Initial skepticism evolves into trust after multiple positive experiences	—	Anticipates a future shift in acceptance of AI's advice, especially without a photo, within a decade
The value of human interaction	—	—	—	Asserts that physical tools and examination are indispensable for accurate diagnostics	Stresses the irreplaceable role of human involvement in making accurate diagnoses

[a]AI: artificial intelligence.

[b]Not applicable.

For the frameworks (Tables 3 and 4), instead of using quotes from focus group participants, we chose to provide summaries of the key findings, following the common practice in thematic framework analysis. By creating summaries, we were able to handle the complex discussion, which involved a record with 7 speakers, more easily.

Interpretations of Thematic Framework Analysis

On the basis of the thematic framework analysis of podiatrists' preferences for AI integration in clinical settings, a clear pattern emerges regarding the desired application and form of transparency in AI systems for triage. Podiatrists prefer using AI in triage, emphasizing the importance of transparency that goes beyond just being open. They are looking for a kind of transparency that involves clinically relevant and detailed information—such as patient history, specific conditions such as the presence of vascular diseases, and exact wound locations—that significantly augment the decision-making process:

> *The added value for me is that the ranking is done by the AI, so photos are prioritised based on urgency.* [Participant 3, male, age group 41-50 years]
>
> *My preference is for triage; an AI can perform it much faster than a human. Well, I believe that AI is capable of making a risk assessment.* [Participant 1, female, age group 41-50 years]
>
> *Key abnormalities must be quickly recognised and addressed, achievable through triage.* [Participant 5, female, age group 61-70 years]

In the domain of triage, skepticism toward patient-reported data as the sole basis for AI-generated advice is evident (eg, medium transparency mock-up 1). The concern centers on the reliability of such data, highlighting a preference for AI systems that incorporate a comprehensive training dataset to inform their assessments. This reflects a broader apprehension regarding the accuracy of AI recommendations based on potentially incomplete or inaccurate patient self-assessments:

> *I'm not entirely convinced by the patients' responses. If they say they have a wound on their feet, then yes, I might believe it, but if they don't have a wound, I'm sceptical.* [Participant 1, female, age group 41-50 years]
>
> *Patients often do not have a clear understanding of complications.* [Participant 3, male, age group 41-50 years]

Podiatrists recognize the potential benefits but are cautious regarding the use of AI for diagnostic applications, particularly those involving the analysis of photographic evidence of foot conditions. This caution is based on the understanding that certain conditions cannot be fully diagnosed through images alone, emphasizing the need for AI systems to supplement rather than replace traditional diagnostic methods:

> *Computers can't replace humans in diagnostics, especially not for tasks requiring nuanced observation, like identifying underlying issues beneath a callus that AI might miss.* [Participant 5, female, age group 61-70 years]
>
> *AI is incapable of making diagnoses. AI cannot handle complex conditions like distinguishing between warm and cold wounds.* [Participant 4, female, age group 51-60 years]
>
> *Even a human can't make a diagnosis just from an image. The foot needs to be examined in person and potentially opened up for an accurate assessment.* [Participant 1, female, age group 41-50 years]

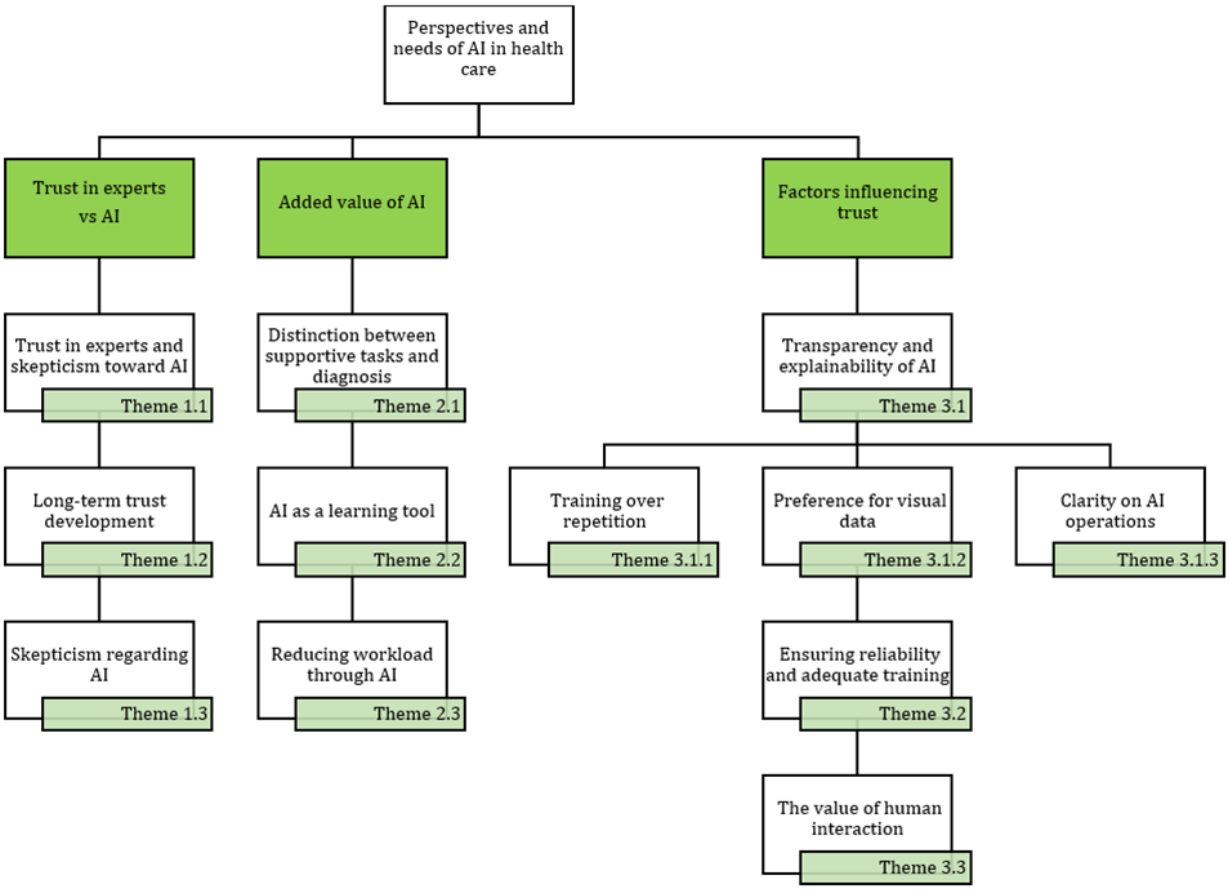
The desired level of transparency in diagnostic AI varies among podiatrists. While some value additional explanatory features, such as annotated images or textual diagnostics, others caution against the potential for such features to overly focus attention on a specific area, possibly leading to the overlooking of important symptoms in other areas. This division highlights the need for a careful balance in using AI; while it is beneficial to receive detailed, actionable insights, it is crucial to ensure that these insights do not overshadow the importance of using one's own professional knowledge and experience in decision-making:

> *Works well for me; the layer aids in understanding the basis of the text.* [Participant 3, male, age group 41-50 years]
>
> *Also, here, the text is okay. I also like the layer, so now I know what the text is referring to.* [Participant 1, female, age group 41-50 years]
>
> *The explanatory text doesn't interest me much; I'll take it on faith what it's based on.* [Participant 4, female, age group 51-60 years]
>
> *And that layer, I can see that for myself, too. Adding a layer can also be risky; I see even more things, and now AI dictates where I should look. A less experienced podiatrist might then focus solely on that.* [Participant 5, female, age group 61-70 years]

In addition to emphasizing transparency in integrating AI into clinical podiatry practice, our analysis has other predefined themes within the framework (Table 4). These themes enrich our understanding of how AI can be effectively applied in triage and diagnostics, emphasizing the need for clear, actionable insights and the importance of trust and data reliability as foundational elements. Reflecting on the preferences expressed by podiatrists, it becomes apparent that alongside transparency and explainability, considerations around data reliability, human interaction, trust development, and trust establishment align with sentiments from the first phase of our study.

However, we have defined 3 additional themes that resonated strongly with the initial input for transparency and explainability. Figure 6 presents a comprehensive overview of all the themes, including the 3 subthemes (ie, 3.1.1: Training Versus Repetition; 3.1.2: Preference for Visual Data; and 3.1.3: Clarity on AI Operations) and their implications for integrating AI into podiatry. The following sections detail these additional subthemes further.

Figure 6. Generated themes and subthemes from phases 1 and 2: perspectives of experienced podiatrists on the FootCheck app and decision-making process. AI: artificial intelligence.



Subtheme 3.1.1: Training Versus Repetition

While participants in phase 1 primarily highlighted concerns among podiatrists regarding data reliability, phase 2 revealed a nuanced understanding of how and when this information should be introduced. This is particularly relevant during the initial training sessions where podiatrists are first taught how to use the AI system. Feedback on mock-up 2 indicated that emphasizing data reliability in every AI recommendation is not always seen as necessary. Some participants supported this viewpoint, while others preferred having this information included. In addition, there was consensus that explanations should offer insights beyond basic knowledge, aiming to augment understanding rather than merely repeat well-known information:

> *In the introduction (training), mention somewhere that the data is based on 5000 photos, but that doesn't need to be in every piece of advice.* [Participant 4, female, age group 51-60 years]
>
> *I expect an AI to be backed by a vast photo database, so it's unnecessary to reiterate that point for me.* [Participant 5, female, age group 61-70 years]
>
> *The text is helpful; it lets me know what the AI is based on.* [Participant 3, male, age group 41-50 years]

Subtheme 3.1.2: Preference for Visual Data

There is a significant inclination toward incorporating visual data alongside algorithmic risk assessments. Participants consider combining photos with textual advice crucial, enhancing their ability to make knowledgeable decisions. The need for visual confirmation to solidify trust in AI-generated advice is highlighted, with participants expressing reservations about relying entirely on text-based recommendations. Although color-coded elements are appreciated for their clarity, the facility to view photos directly via pop-ups is deemed essential for building confidence in the advice provided:

> *The colour coding in mock-up 1 is helpful, but I'd prefer to see the photos appear as pop-ups.* [Participant 4, male, age group 41-50 years]
>
> *Just text, without a photo, doesn't inspire my trust.* [Participant 3, female, age group 51-60 years]
>
> *It seems like a small ask to quickly review the photo for a better trust in the advice.* [Participant 4, female, age group 51-60 years]

Subtheme 3.1.3: Clarity on AI Operations

Podiatrists highlight the necessity for transparency regarding the process behind AI-generated advice, which is essential for building trust and enabling users to understand when and how to rely on these recommendations. Opinions vary on the best way to present this information, with some finding objective data in photos and layered visuals helpful for understanding the advice's basis, as seen in mock-up 2. Others question the interpretability of numerical data, such as "73%," without seeing the patient in person:

> *Now, with a photo and layer, the information is more objective; you know what the text is about.* [Participant 3, male, age group 41-50 years]
>
> *I appreciate it, as then I know what it's based on, mock-up 2.* [Participant 1, female, age group 51-60 years]

Discussion

AI in Supportive Roles Versus Diagnostic Tasks

Our study aimed to discern the multifaceted roles AI could assume in health care, specifically focusing on how AI can be effectively integrated into practice in a way that ensures health care professionals' acceptance and trust in its advice. The findings from our studies suggest a preference for deploying AI in supportive roles. For instance, AI's use in triaging processes is evident; by evaluating submitted photos for urgency, AI can advise health care providers in prioritizing patient consultations. This enables a more efficient allocation of resources, ensuring that patients requiring urgent care receive it promptly, while less-critical cases are appropriately prioritized. Such applications align with and enrich suggestions from the literature, including studies by Verma et al [10] and Longoni et al [5], suggesting the supportive use of AI to foster trust within health care settings. However, a contrasting perspective is presented in another experiment by Longoni et al [5], where preference leans toward human expertise over automated systems for triage tasks. This discrepancy from our findings may stem from differing viewpoints, with the study results of Longoni et al [5] being based on patient perspectives, whereas our study was performed among health care providers.

Another critical insight from our study highlights health care professionals' hesitancy to embrace AI for diagnostic purposes, a sentiment that aligns with previous research findings [5,7,9,10,18]. While there is acknowledgment of AI's proficiency in critical functions, such as making diagnoses, with results that match or even surpass human expertise, there remains skepticism among participants about AI's current proficiency in performing these tasks accurately. The observed reluctance may stem from unfamiliarity with the full scope of AI's capabilities, as evidenced by the fact that particularly participants without prior AI experience expressed reservations. Moreover, the conversations uncovered that identifying conditions such as diabetic foot complications poses distinct obstacles for AI. The participants pointed out that superficial evaluations, such as analyzing diabetic wounds via photographs, could be inadequate without the physical removal of calluses to reveal the actual problem, an action AI is incapable of without a physical examination. This underlines the importance of direct human interaction in specific diagnostic processes, highlighting limitations in the potential application of AI.

Algorithm Aversion Among Podiatrists

The observed phenomenon of algorithm aversion among podiatrists, particularly their preference for human expertise over AI-generated recommendations, reflects a trend noted across various fields where AI competes with human judgment [26-29]. For example, Önkal et al [30] demonstrated a similar inclination among forecasters to favor expert opinions in forecasting decisions. A notable factor contributing to this aversion is the concern over mistakes made by AI systems [26], with our podiatrists voicing worries about the grave repercussions of such errors, including the risk of limb amputation. Although the significance of decisions ranges from health care to forecasting tasks in logistics, the underlying dread of AI-induced errors remains pervasive.

However, in our studies, this aversion was primarily directed at diagnostic tasks rather than supportive functions. Participants expressed confidence in the expertise and experience of their colleagues but also showed openness to considering AI-generated advice for tasks such as triage. This nuanced view suggests a complex relationship between podiatrists and AI, where trust may be contingent on the specific task at hand and the perceived reliability of AI in enhancing rather than replacing human judgment.

Transparency and Explainability

The discussions surrounding transparency and explainability in our studies mirrored the diversity of perspectives found in the broader literature on these subjects [12,14,19]. The need for explanation and transparency depends significantly on podiatrists' individual preferences. Our study provides 3 key insights regarding the implementation of explainability and transparency in AI systems.

First, precise and valuable explanations that avoid redundancy are essential for enhancing user understanding and trust in AI. Second, visual aids alongside textual advice can boost confidence in AI recommendations more than text alone, by making the motivation behind complex decisions more accessible. Finally, it is important to remind users to balance AI advice with their professional judgment, highlighting the crucial role of human oversight in AI decision-making.

However, these discussions were particularly animated regarding the scenario involving AI-generated diagnoses, where participants displayed heightened criticality toward the information received from AI. This suggests that individuals are particularly cautious about relying on AI for diagnostic tasks, scrutinizing the transparency and explainability of such systems with greater intensity. Interestingly, these issues were less extensively debated in contexts involving AI for supportive tasks. This observation highlights an area for future research.

Development of Trust Over Time

Most participants indicated a higher level of trust in their colleagues than in AI systems, suggesting that trust in AI might evolve positively with increased exposure and evidence of AI's efficacy. However, this remains speculative, as our participants have not extensively experimented with AI in their professional practice. The notion that repeated interactions with AI could enhance trust aligns with existing literature [31,32]. This area, highlighting the dynamic interplay between growing familiarity with AI and the shifting landscape of trust and reliance, remains underexplored in our studies and represents a fruitful avenue for further research.

Strengths and Limitations

Our research is strengthened by several methodological approaches that enhance the credibility of our findings. The employment of 2 independent researchers for parallel data coding in both studies significantly contributed to the reliability of our results, as evidenced by the substantial agreement in their coding outcomes [33].

The relatively small sample sizes of 9 podiatrists in the first phase and 5 in the second phase may seem to be a limitation for generating general conclusions. However, qualitative research often reaches a point of data saturation, where small samples are already sufficient to uncover dependable themes. In our study, we observed that no new codes or themes emerged after 7 interviews in the first phase, indicating that data saturation had been reached. This aligns with the findings of Guest et al [34], who noted that most codes are identified within the first 6 interviews. Guest et al [35] also supported this by providing a systematic method for assessing saturation, further validating our sample size.

The significant expertise of the podiatrists involved, that is, all specialists in the area considered, adds depth to our insights and strengthens the credibility of our findings. In addition, the second phase, with mostly new participants, confirmed the results of phase 1. By focusing on how AI might be more accepted in supportive tasks compared with diagnostic tasks, we were able to further validate the earlier findings and ensure that no new themes emerged, indicating that saturation had been reached.

Conclusions

In conclusion, the findings reveal podiatrists' complex attitudes toward AI in health care, demonstrating openness to AI assistance for supportive tasks while being cautionary regarding its application for diagnostics. This underscores the necessity for developers and policy makers to prioritize trust-building strategies, starting with opportunities that allow podiatrists to experience firsthand how AI can support their role. Such initial experiences could clarify AI's benefits and capabilities to users, reduce perceived threats, and facilitate a smoother transition toward acceptance. With a foundation of trust established through the application of AI for supportive tasks, podiatrists may become more receptive to exploring AI's potential for diagnostic tasks. Gradually expanding AI's role from foundational tasks, including decision support, to areas requiring deeper trust, such as diagnostics, could enhance acceptance and integration of AI technologies in podiatric practice.

Acknowledgments

The authors extend their heartfelt gratitude to all individuals who generously shared their insights and experiences, making this research possible. Special thanks are extended to Ruben Gloudemans for his invaluable assistance with conducting interviews, engaging in thorough analyses and participating in wide-ranging discussions that significantly enriched this work. The authors appreciate RondOm Podiatrists for providing access to the FootCheck app, an essential resource for our study. This work is part of the project Vertrouwen in Algoritmes (project number 023.017.109), under the auspices of the Promotiebeurs voor Leraren program, generously funded by the Dutch Research Council (NWO). Additionally, this study was part of the Technology Support for Diabetes project, which is subsidized by SIA RAAK EXZ.EXZ.01.004. The research efforts of the authors were greatly supported by these contributions, for which they are profoundly grateful.

Data Availability

The corresponding author will make the datasets supporting these studies available upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide used in phase 1.
[DOCX File , 20 KB - humanfactors_v12i1e59010_app1.docx]

Multimedia Appendix 2
Mock-ups used in phase 2.
[DOCX File , 1981 KB - humanfactors_v12i1e59010_app2.docx]

Multimedia Appendix 3
Interview guide used in phase 2.
[DOCX File , 16 KB - humanfactors_v12i1e59010_app3.docx]

References

1. Bazazeh D, Shubair R. Comparative study of machine learning algorithms for breast cancer detection and diagnosis. In: Proceedings of the 5th International Conference on Electronic Devices, Systems and Applications. 2017 Presented at: ICEDSA '16; December 6-8, 2016; Ras Al Khaimah, United Arab Emirates p. 1-4 URL: https://ieeexplore.ieee.org/document/7818560 [doi: 10.1109/icedsa.2016.7818560]
2. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
3. Eriksen AV, Möller S, Ryg J. Use of GPT-4 to diagnose complex clinical cases. NEJM AI 2024 Jan 11;1(1):24. [doi: 10.1056/aip2300031]
4. Meng Y, Sun J, Qu N, Zhang G, Yu T, Piao H. Application of radiomics for personalized treatment of cancer patients. Cancer Manag Res 2019;11:10851-10858 [FREE Full text] [doi: 10.2147/CMAR.S232473] [Medline: 31920394]
5. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res 2019;46(4):629-650. [doi: 10.1093/jcr/ucz013]
6. Promberger M, Baron J. Do patients trust computers? J Behav Decis Mak 2006 Nov 24;19(5):455-468. [doi: 10.1002/bdm.542]
7. Seitz L, Bekmeier-Feuerhahn S, Gohil K. Can we trust a chatbot like a physician? A qualitative study on understanding the emergence of trust toward diagnostic chatbots. Int J Hum Comput Stud 2022 Sep;165:102848. [doi: 10.1016/j.ijhcs.2022.102848]
8. Naik N, Hameed BM, Shetty DK, Swain D, Shah M, Paul R, et al. Legal and ethical consideration in artificial intelligence in healthcare: who takes responsibility? Front Surg 2022 Mar 14;9:862322 [FREE Full text] [doi: 10.3389/fsurg.2022.862322] [Medline: 35360424]
9. Juravle G, Boudouraki A, Terziyska M, Rezlescu C. Trust in artificial intelligence for medical diagnoses. Prog Brain Res 2020;253:263-282. [doi: 10.1016/bs.pbr.2020.06.006] [Medline: 32771128]
10. Verma H, Mlynar J, Schaer R, Reichenbach J, Jreige M, Prior J, et al. Rethinking the role of AI with physicians in oncology: revealing perspectives from clinical and research workflows. In: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023 Presented at: CHI '23; April 23-28, 2023; Hamburg, Germany p. 11-52 URL: https://dl.acm.org/doi/fullHtml/10.1145/3544548.3581506 [doi: 10.1145/3544548]
11. Logg JM, Minson JA, Moore DA. Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process 2019 Mar;151:90-103. [doi: 10.1016/j.obhdp.2018.12.005]
12. Naiseh M, Al-Thani D, Jiang N, Ali R. How the different explanation classes impact trust calibration: the case of clinical decision support systems. Int J Hum Comput Stud 2023 Jan;169:102941. [doi: 10.1016/j.ijhcs.2022.102941]
13. Panigutti C, Beretta A, Giannotti F, Pedreschi D. Understanding the impact of explanations on advice-taking: a user study for AI-based clinical decision support systems. In: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems. 2022 Presented at: CHI '22; April 29-May 5, 2022; New Orleans, LA p. 1-9 URL: https://dl.acm.org/doi/10.1145/3491102.3502104 [doi: 10.1145/3491102.3502104]
14. Alam L, Mueller S. Examining the effect of explanation on satisfaction and trust in AI diagnostic systems. BMC Med Inform Decis Mak 2021 Jun 03;21(1):178. [doi: 10.1186/s12911-021-01542-6] [Medline: 34082719]
15. Tucci V, Saary J, Doyle TE. Factors influencing trust in medical artificial intelligence for healthcare professionals: a narrative review. J Med Artif Intell 2022 Mar;5:4. [doi: 10.21037/jmai-21-25]
16. Nagendran M, Festor P, Komorowski M, Gordon AC, Faisal AA. Quantifying the impact of AI recommendations with explanations on prescription decision making. NPJ Digit Med 2023 Nov 07;6(1):206. [doi: 10.1038/s41746-023-00955-z] [Medline: 37935953]
17. Prahl A, Van Swol L. Understanding algorithm aversion: when is advice from automation discounted? J Forecast 2017 Mar 24;36(6):691-702. [doi: 10.1002/for.2464]
18. Hallowell N, Badger S, Sauerbrei A, Nellåker C, Kerasidou A. "I don't think people are ready to trust these algorithms at face value": trust and the use of machine learning algorithms in the diagnosis of rare disease. BMC Med Ethics 2022 Nov 16;23(1):112 [FREE Full text] [doi: 10.1186/s12910-022-00842-4] [Medline: 36384545]
19. Schwartz JM, George M, Rossetti SC, Dykes PC, Minshall SR, Lucas E, et al. Factors influencing clinician trust in predictive clinical decision support systems for in-hospital deterioration: qualitative descriptive study. JMIR Hum Factors 2022 May 12;9(2):e33960 [FREE Full text] [doi: 10.2196/33960] [Medline: 35550304]
20. Lee JT, Hsieh CC, Lin CH, Lin YJ, Kao CY. Prediction of hospitalization using artificial intelligence for urgent patients in the emergency department. Sci Rep 2021 Sep 30;11(1):19472 [FREE Full text] [doi: 10.1038/s41598-021-98961-2] [Medline: 34593930]
21. Crampton NH. Ambient virtual scribes: Mutuo Health's AutoScribe as a case study of artificial intelligence-based technology. Healthc Manage Forum 2020 Jan;33(1):34-38. [doi: 10.1177/0840470419872775] [Medline: 31522566]
22. Berings M, Arts M, Pelders D, Sirag S, Brankaert R, Nuijten E E. Technologie helpt diabetische voetenzorg vooruit. ICT Health 2022 Apr 14;46(2):82-85 [FREE Full text]
23. Transform your audio and video to text and subtitles. Amberscript Global. URL: https://www.amberscript.com/en/ [accessed 2024-04-29]

24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. Collaço N, Wagland R, Alexis O, Gavin A, Glaser A, Watson EK. Using the framework method for the analysis of qualitative dyadic data in health research. Qual Health Res 2021 May 13;31(8):1555-1564. [doi: 10.1177/10497323211011599]
26. Dietvorst BJ, Simmons JP, Massey C. Overcoming algorithm aversion: people will use imperfect algorithms if they can (even slightly) modify them. Manag Sci 2018 Mar;64(3):1155-1170. [doi: 10.1287/mnsc.2016.2643]
27. Mahmud H, Islam AK, Ahmed SI, Smolander K. What influences algorithmic decision-making? A systematic literature review on algorithm aversion. Technol Forecast Soc Change 2022 Feb;175:121390. [doi: 10.1016/j.techfore.2021.121390]
28. Castelo N, Bos MW, Lehmann DR. Task-dependent algorithm aversion. J Mark Res 2019 Jul 15;56(5):809-825. [doi: 10.1177/0022243719851788]
29. Burton JW, Stein MK, Jensen TB. A systematic review of algorithm aversion in augmented decision making. Behav Dis Mark 2019 Oct 23;33(2):220-239. [doi: 10.1002/bdm.2155]
30. Önkal D, Goodwin P, Thomson M, Gönül S, Pollock A. The relative influence of advice from human experts and statistical methods on forecast adjustments. Behav Decis Mark 2009 Feb 09;22(4):390-409. [doi: 10.1002/bdm.637]
31. Lim JS, O'Connor M. Judgmental forecasting with interactive forecasting support systems. Decis Support Syst 1996 Apr;16(4):339-357. [doi: 10.1016/0167-9236(95)00009-7]
32. Whitecotton SM. The effects of experience and a decision aid on the slope, scatter, and bias of earnings forecasts. Organ Behav Hum Decis Process 1996 Apr;66(1):111-121. [doi: 10.1006/obhd.1996.0042]
33. Malterud K. Qualitative research: standards, challenges, and guidelines. Lancet 2001 Aug 11;358(9280):483-488. [doi: 10.1016/S0140-6736(01)05627-6] [Medline: 11513933]
34. Guest G, Bunce A, Johnson L. How many interviews are enough?: an experiment with data saturation and variability. Field Methods 2006 Feb 01;18(1):59-82. [doi: 10.1177/1525822X05279903]
35. Guest G, Namey E, Chen M. A simple method to assess and report thematic saturation in qualitative research. PLoS One 2020 May 5;15(5):e0232076 [FREE Full text] [doi: 10.1371/journal.pone.0232076] [Medline: 32369511]

Abbreviations

AI: artificial intelligence
XAI: explainable artificial intelligence

Edited by E Borycki, A Kushniruk; submitted 30.03.24; peer-reviewed by S Sadler; comments to author 08.07.24; revised version received 18.10.24; accepted 29.12.24; published 21.02.25.

Please cite as:
Tahtali MA, Snijders CCP, Dirne CWGM, Le Blanc PM
Prioritizing Trust in Podiatrists' Preference for AI in Supportive Roles Over Diagnostic Roles in Health Care: Qualitative Interview and Focus Group Study
JMIR Hum Factors 2025;12:e59010
URL: https://humanfactors.jmir.org/2025/1/e59010
doi:10.2196/59010
PMID:

Research Letter

Older Adults' Experiences With an Online Survey

Kristie Rebecca Weir[1,2], PhD; Yehya Maitah[3], BS; Sarah E Vordenberg[4], PharmD, MPH

[1]Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia

[2]Institute of Primary Health Care (BIHAM), University of Bern, Bern, Switzerland

[3]Albany Medical College, Albany, NY, United States

[4]University of Michigan College of Pharmacy, Department of Clinical Pharmacy, Ann Arbor, MI, United States

Corresponding Author:
Kristie Rebecca Weir, PhD
Sydney School of Public Health
Faculty of Medicine and Health
The University of Sydney
City Road
Sydney, NSW 2006
Australia
Phone: 61 293512222
Email: kristie.weir@sydney.edu.au

Abstract

The study explored older adults' perceptions after participating in an online survey about medication decisions, finding that approximately 80% of participants provided positive feedback about the research methodology and their experience.

(JMIR Hum Factors 2025;12:e65684) doi:10.2196/65684

KEYWORDS

older adults; gerontology; geriatric; older; aging; online; internet; survey; questionnaire; research engagement; engagement; study subject; participant; medication decisions

Introduction

Older adults are underrepresented in research, often due to age-related biases and stringent exclusion criteria; this limits generalizability and leaves knowledge gaps [1].

The increasing prevalence of online research has the potential to increase older adult participation, given the rising internet use among this demographic [2]. While research has primarily focused on improving recruitment of older adults, an understanding of their experiences of online research is needed to effectively engage older people [3]. While most surveys include some form of piloting, they rarely capture or share participants' perceptions of the research process. We sought to explore older adults' perceptions after participating in an online experimental survey, given the potential high accessibility of this type of research.

Methods

Overview

We previously reported the main findings from a vignette-based online experiment conducted among adults aged 65 years and older from Australia, the Netherlands, the United Kingdom, and the United States [4]. Participants were recruited for this 15-minute survey through a panel of internet users administered by Qualtrics Research Panels. The study focused on contextual factors influencing how older adults (mean age 71.5, SD 5.1 years) think about deciding to stop a cardiovascular disease medication. In this study, we conducted a content analysis of free-text comments (Textbox 1). The final analysis framework included 17 codes that were subsequently consolidated into 14 themes. Descriptive statistics were used to assess the frequency of each theme.

Textbox 1. Summary of methodology for the content analysis.

1. A vignette-based online experiment was conducted among adults aged 65 years and older from Australia, the Netherlands, the United Kingdom, and the United States using sampling quotas to ensure balanced representation by country and gender. This study was registered at ClinicalTrials.gov (NCT04676282). At the end of the survey, participants were provided the opportunity to leave "any comments about this study."
2. A content analysis was conducted among participants who received the vignette about a hypothetical patient, Mrs EF. Participants from the Netherlands were excluded, as the study team for the content analysis was not fluent in Dutch.
3. Study authors read through the comments to inductively generate codes.
4. Two investigators (SEV and YM) independently coded comments, resulting in >80% agreement.
5. All discrepancies in codes were discussed until consensus was reached.

Ethical Considerations

The online experiment was registered at ClinicalTrials.gov (NCT04676282) and was deemed exempt by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board (HUM00183129); by extension, a waiver of consent was granted. All data were collected anonymously. Participants were compensated based on the terms of their panel agreement.

Results

Participants (N=1789) most frequently did not provide any feedback in the free-text comment field (n=784, 43.8%) or wrote that they had no comments (n=487, 27.2%). Three participants (0.2%) gave unclear statements.

Participants' comments (n=515, 28.8%) were primarily positive (415/515, 80.6%), such as that the study was interesting (116/515, 22.5%). Themes and representative quotes are reported in Table 1. Participants said it made them think about their health (80/515, 15.5%) and some participants shared further health information about the study topic (48/515, 9.3%). Participants provided feedback on how to improve the study, categorized as question-specific comments (21/515, 4.1%), general suggestions (12/515, 2.3%), or country-specific comments (3/515, 0.6%). Few participants (16/515, 3.1%) provided negative feedback about the survey.

Table 1. Older adults' feedback about an online survey using a hypothetical vignette by theme, with representative quotes among respondents who provided feedback (n=515).

Themes	Representative quotes (participant code)	Participants, n (%)
Positive feedback (n=415, 80.6%)		
Interesting	• "An interesting study" (919) • "This was an interesting survey to complete" (2200)	116 (22.5)
Thought provoking	• "Interesting and insightful. Made me think a little more about how best to manage my health" (1872) • "Very good survey to ponder thoughts and beliefs" (481)	80 (15.5)
Positive feedback	• "Great stuff. Keep up the good work" (2574) • "Very good" (764)	67 (13)
Thanks for opportunity to participate	• "Thanks" (42)	43 (8.3)
Positive feedback about survey questions or structure	• "Love the format, so easy to see, follow, and understand" (1904) • "Very good questions regarding whether or not to stop a medication when you've been on it for an extended time" (2026)	35 (6.8)
Enjoyed taking survey	• "Enjoyed it" (328) • "Love doing your studies" (287)	34 (6.6)
Unusual study design	• "Very different enjoyed the variety" (540) • "Unusual but interesting" (1853)	19 (3.7)
Interested in results	• "Interested in purpose of results" (1863) • "Very interesting! Now I need to see the final results" (3481)	15 (2.9)
Interested in future studies	• "Great study, I'd do more" (4815) • "Need more like this" (2516)	6 (1.2)
Neutral feedback (n=48, 9.3%)		
Shared personal experience	• "I am afraid of going to see the dr for worries about my health" (1526) • "Thanks; my prescription drugs fall under the Federal Government program where they are funded nearly 100%" (2645)	48 (9.3)
Negative feedback or suggestions for improvement (n=52, 10.1%)		
Question-specific feedback	• "The percentage charts were confusing" (2403) • "One question said check all that apply but only one was allowed" (4279)	21 (4.1)
Negative feedback	• "Boring" (1409) • "Too many generalizations" (1102)	16 (3.1)
General suggestions for improvement	• "Consider adding a progress bar to the survey" (2027) • "Good study could be shorter" (2433	12 (2.3)
Country-specific feedback	• "Some questions are designed for the USA" (1471) • "Prescription insurance? In UK if you are over 65 prescriptions are free" (4075)	3 (0.6)

Discussion

Among older adults who provided feedback about their experiences completing an online survey, approximately 80% (415/515) of the comments were positive. Our findings signal general acceptability of the methodology, and we have implemented the practical feedback to improve our online surveys. We have become more mindful of the survey length, selected straightforward question types, and have conducted pilot testing in all target countries to ensure that questions are appropriate for all participants.

Our study had several limitations. First, less than one-third of the study participants provided any comments. Second, participants were asked if they had any comments, as opposed to more specific questions about their experience taking the survey. Finally, we acknowledge that we coded a single primary theme per comment given the short statements that were provided.

With an aging population who may spend many years in retirement, participating in research can offer benefits such as reducing social isolation and loneliness, fostering a sense of

purpose, and providing mental stimulation, and it may provide monetary incentives [5]. While our online study lacked the social benefit of in-person interaction typical of traditional research, participants reported enjoying the survey; they found it prompted self-reflection on their health, and they expressed interest in the study's outcomes and future research.

Online research is becoming more prevalent; therefore, it is important to make sure this methodology is inclusive of older adults [5]. More than half of adults older than 65 years use the internet, yet they are the least likely to have a home computer [6]. While some researchers attribute this limited use to age-related functional decline, others argue that the main barriers are negative attitudes such as fear, anxiety, and low motivation—barriers that are modifiable. Anxiety about using the internet and digital technologies often leads to self-imposed limitations and low confidence, with older adults frequently underestimating their knowledge and abilities compared to younger users [6]. This underscores the importance of studies like ours that highlight the positive experiences of older adults in online research.

Acknowledgments

This study was funded by award R24AG064025 from the US Deprescribing Research Network via the Northern California Institute for Research and Education through the National Institute on Aging; a Swiss Government Excellence Scholarship (KRW), the Swiss National Science Foundation Scientific Exchanges (grant IZSEZ0_213475/1; KRW), and an Emerging Leader Research Fellowship (2017295) from the Australian National Health and Medical Research Council (KRW). The funders had no role in the design or conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. KRW and SEV had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Authors' Contributions

Concept and design: KRW, SEV
Acquisition, analysis, or interpretation of data: all authors
Drafting of the manuscript: all authors
Critical review of the manuscript for important intellectual content: all authors
Statistical analysis: SEV
Obtained funding: SEV
Administrative, technical, or material support: KRW
Supervision: KRW, SEV

Conflicts of Interest

KRW reports receiving grant funding from the Swiss National Science Foundation (SNSF), the National Health and Medical Research Council, and the Swiss Confederation during the conduct of the study. SEV reports receiving grant funding from the US Deprescribing Research Network during the conduct of the study.

References

1. Markham SC, McNab J, O'Loughlin K, Clemson L. International responses addressing the under-representation of older people in clinical research. Australas J Ageing 2023 Dec;42(4):762-768. [doi: 10.1111/ajag.13234] [Medline: 37724905]
2. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Research Center. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2024-08-12]
3. Remillard ML, Mazor KM, Cutrona SL, Gurwitz JH, Tjia J. Systematic review of the use of online questionnaires of older adults. J Am Geriatr Soc 2014 Apr;62(4):696-705 [FREE Full text] [doi: 10.1111/jgs.12747] [Medline: 24635138]
4. Weir KR, Vordenberg SE, Scherer AM, Jansen J, Schoenborn N, Todd A. Exploring different contexts of statin deprescribing: a vignette-based experiment with older adults across four countries. J Gen Intern Med 2024 Jul;39(9):1773-1776 [FREE Full text] [doi: 10.1007/s11606-024-08698-7] [Medline: 38514582]
5. Lindeman D. Improving the independence of older adults through technology: directions for public policy. Public Policy Aging Rep 2017;27(2):49-52. [doi: 10.1093/PPAR/PRX011]
6. Vroman KG, Arthanat S, Lysack C. "Who over 65 is online?" Older adults' dispositions toward information communication technology. Comput Human Behav 2015 Feb;43:156-166. [doi: 10.1016/j.chb.2014.10.018]

Edited by A Kushniruk, E Borycki; submitted 22.08.24; peer-reviewed by J Rodger, S Serhal, A Vassallo; comments to author 19.12.24; revised version received 19.12.24; accepted 20.12.24; published 08.01.25.

Please cite as:
Weir KR, Maitah Y, Vordenberg SE
Older Adults' Experiences With an Online Survey
JMIR Hum Factors 2025;12:e65684
URL: https://humanfactors.jmir.org/2025/1/e65684
doi:10.2196/65684
PMID:

Factors Associated With the Intention to Use mHealth Among Thai Middle-Aged Adults and Older Adults: Cross-Sectional Study

Nida Buawangpong[1], MD; Wachiranun Sirikul[2,3,4,5], MSc, MD; Penprapa Siviroj[2,4], DrPH

[1]Department of Family Medicine, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[2]Department of Community Medicine, Faculty of Medicine, Chiang Mai University, 110, Intrawarorot road, Sriphum, Meaung, Chiang Mai, Thailand

[3]Center of Data Analytics and Knowledge Synthesis for Health Care, Chiang Mai University, Chiang Mai, Thailand

[4]Environmental and Occupational Medicine Excellence Center, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

[5]Department of Biomedical Informatics and Clinical Epidemiology, Faculty of Medicine, Chiang Mai University, Chiang Mai, Thailand

Corresponding Author:
Wachiranun Sirikul, MSc, MD
Department of Community Medicine, Faculty of Medicine, Chiang Mai University, 110, Intrawarorot road, Sriphum, Meaung, Chiang Mai, Thailand

Abstract

Background: Mobile health care (mHealth) apps are emerging worldwide as a vital component of internet health care, but there are issues, especially among older adults.

Objective: We aim to investigate the factors influencing the intention to use (ITU) mHealth apps, focusing on those with and without prior mHealth experience.

Methods: A cross-sectional study conducted from August 2022 to July 2023 included Thai citizens aged 45 years or older. Self-reported questionnaires collected data on sociodemographic information, health conditions, smartphone or tablet ownership, and mHealth usage experience. The Thai mHealth Senior Technology Acceptance Model questionnaires with a 10-point Likert scale evaluated mHealth acceptance. A multivariable logistic regression analysis, adjusted for age, gender, education, income, and living area, was performed for 2 subgroups: those who used ITU mHealth apps and those who did not.

Results: Of 1100 participants, 537 (48.8%) intended to use mHealth apps, while 563 (51.2%) did not. The ITU group had a younger average age, higher education levels, higher income, and fewer underlying diseases compared to those who did not intend to use mHealth apps. For those who had never used mHealth apps, having a smartphone was strongly associated with higher odds of ITU (adjusted odds ratio 2.81, 95% CI 1.6 to 4.93; P<.001), while having any underlying disease was associated with lower odds of ITU (adjusted odds ratio 0.63, 95% CI 0.42 to 0.97; P=.034). Higher acceptance levels, characterized by a positive attitude toward mHealth and lower fear of making mistakes, were also associated with higher ITU. For those with prior mHealth experience, acceptance in areas such as perceived ease of use, gerontechnology anxiety, and facilitating conditions was significantly associated with ITU.

Conclusions: Among inexperienced users, a positive attitude toward mHealth significantly enhanced ITU. Conversely, having an underlying disease decreased ITU, indicating a need for tailored mHealth apps. For experienced users, acceptance levels in areas such as ease of use and gerontechnology anxiety were crucial. Future research should explore specific mHealth apps for more targeted insights.

(JMIR Hum Factors 2025;12:e63607) doi:10.2196/63607

KEYWORDS

mHealth; mobile healthcare; older adults; elderly; aging; questionnaire; smartphone; mHealth usage; intention to use

Introduction

Mobile health care (mHealth) apps, a vital component of internet health care, are emerging worldwide. These apps have attracted a wide range of health care services and are proven effective in solving health problems. The functions of mHealth are diverse, including patient monitoring, diagnosis, personal care, psychological health, educational apps, and social networking [1]. There is growing evidence of the health benefits of mHealth. Mobile devices, with their integrated sensors and features, help health care professionals treat patients with continuous connectivity. These apps are useful for collecting data related to physical activity, human body images, and other health care aspects [2,3]. For example, using mHealth in chronic disease management has shown improvements in symptoms and reduced hospitalizations for patients with asthma, chronic obstructive pulmonary disease, heart failure, and diabetes [4].

However, there are several issues with current mHealth apps, particularly among older adults. Studies show that only 60% of older people intend to use mHealth [5]. Older adults may have physiological changes according to aging such as visual and hearing decline. They may also be unfamiliar with technology and face difficulty learning new skills [6]. Some users experience technical problems with their smartphones when using these apps. Additionally, public awareness of mHealth apps is low, and their usability is not as good as expected [7].

Understanding the factors that affect the intention to use (ITU) mHealth among people with and without prior experience is crucial for enhancing its future usage [8]. Therefore, the primary objective of our study was to investigate the factors associated with ITU, focusing on 2 specific subgroups: those with prior experience using mHealth apps and those without. This study also examined information regarding device ownership, experience with mobile apps and mHealth, and technology acceptance. These findings aimed to gain a better understanding of mHealth usage and identify potential opportunities for implementing mHealth in a community-dwelling population, especially among older adults.

Methods

Study Design

We conducted the cross-sectional study from August 2022 to July 2023 using a nationwide web-based survey and a community survey. The online survey was disseminated through various social media platforms, such as department websites, Facebook, Line, Twitter, and Instagram. The investigator teams, which included medical students and health care personnels from primary care units across 10 subdistricts in Chiang Mai province, distributed the community survey. The respondents to both the online and community surveys used the REDCap (Research Electronic Data Capture; Vanderbilt University) survey platform to self-complete the questionnaires. The STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement guided this study's reporting [9].

Ethical Considerations

The institutional review board of the Faculty of Medicine, Chiang Mai University approved this study's ethical consideration (COM-2565 - 09079). Before participating in this survey, all respondents provided their informed consent in accordance with the screening questionnaire and study information page. Participants were compensated with an incentive of 100 Thai Baht (3 USD) for completing the questionnaires. No identification data were recorded, and respondents were permitted to remain anonymous during the online survey. The community survey used the identification data of eligible participants exclusively for recruitment purposes within each target area. These data were not documented in either the survey form or the study database.

Participants

Study participants were middle-aged Thai citizens aged 45 - 60 years or older at the time of the survey [10,11]. Our study's inclusion criteria required participants to be able to read and communicate in Thai and to have no underlying conditions or diseases that would hinder their ability to complete the survey or use mHealth apps (eg, dementia, active psychological problems, or severe visual impairments). This study excluded respondents who did not complete the survey, spent less than 2 minutes on it, or spent more than 60 minutes on it.

Data Collection

Participant Characteristics

We used self-reported questionnaires to collect data on participant characteristics, including sociodemographic data, underlying health conditions, owning a smartphone or tablet, and experience using the devices and mHealth apps. The initial section of the questionnaires included information regarding mHealth to ensure that the participants comprehended its definition, related concepts, and intended use.

The sociodemographic data included age, gender, marital status (single, married, and separate, divorced, or widowed), living status (alone, with family, and with others), living areas (urban, suburban, and rural), education levels (no education, primary, secondary, high school, vocational training, preuniversity, bachelor's degree, and master's degree), and income per month (<10,000 in Thai Baht, US $ 274; 10,001 - 30,000 in Thai Baht, US $275-$819; and >30,001 in Thai Baht, US $820). The questions inquiring about underlying medical conditions included hypertension, dyslipidemia, diabetes mellitus, chronic renal disease, visual impairments, and hearing impairments. We also gathered data on other related variables, such as wearing glasses or contact lenses, using hearing aids, and a number of current medications.

Experience Using Mobile Apps and mHealth

mHealth apps, or mobile health apps, refer to the practice of medicine and public health supported by mobile devices. The term "mHealth apps" encompasses the use of mobile communication devices such as smartphones and tablets for health care purposes [12,13]. For information on owning and experience using the devices and mHealth, questionnaires asked for smartphone or tablet usage experience (year), owning smartphones or tablets, internet usage experience (year), and previously used mobile apps and mHealth apps.

mHealth Acceptance and ITU

The acceptance of mHealth was evaluated using the Thai mHealth Senior Technology Acceptance Model (STAM) questionnaires. This instrument has been adapted from the 38-item STAM questionnaire [14] and validated for accessing mHealth acceptance in a Thai context. According to the reported psychometric analysis, the Thai mHealth STAM demonstrated satisfactory psychometric properties in terms of validity and reliability. We used instrument items in the following domains as the potential associated factors of an ITU mHealth: attitude toward using, perceived usefulness, perceived ease of use, perceived barriers, gerontechnology anxiety, and facilitating condition. The structure of the Thai mHealth STAM was a 10-point Likert scale. The highest point indicated a high level of acceptance. The participant's ITU mHealth was assessed by the structural question "If there are available mHealth apps for you, do you want to use them? (Yes/No)."

Sample Size

To derive statistics that represent the parameters for the target populations, a minimum sample size of 500 total participants is required. This minimum sample size can accurately detect low to large effect sizes, as suggested by Bujang et al [15]. Based on our primary objective, it aimed to explore the associations between ITU and predetermined variables, including characteristics, ownership, and experience using devices and mHealth, and mHealth acceptance, in 2 specified subgroups of participants: those with experience using mHealth apps and those without. To explore unbiased effect estimates, 18 independent variables were analyzed with adjustments for 5 confounding variables. The sample size calculation, based on the formula n=100+(10 - 50 event per variable)*i*, where *i* refers to the number of independent variables, was conducted to ensure our effect estimates from a multivariable logistic regression had sufficient precision to yield medium to large effect estimates in the subgroup analysis [15]. With 23 independent variables, a minimum sample size of 330 participants per subgroup was needed according to the given formula with 10 per independent variables.

Statistical Analysis

For descriptive analysis, we presented categorical data using frequency and percentage. We described continuous data using a mean with SD, a median with a range (minimum-maximum), or an IQR, as appropriate. The comparison in characteristics, owning and experience using the devices and mHealth, and mHealth acceptance between participants who had ITU and those who did not was examined using an independent 2-tailed *t* test or Wilcoxon rank sum test for continuous data and Fisher exact test for categorical data.

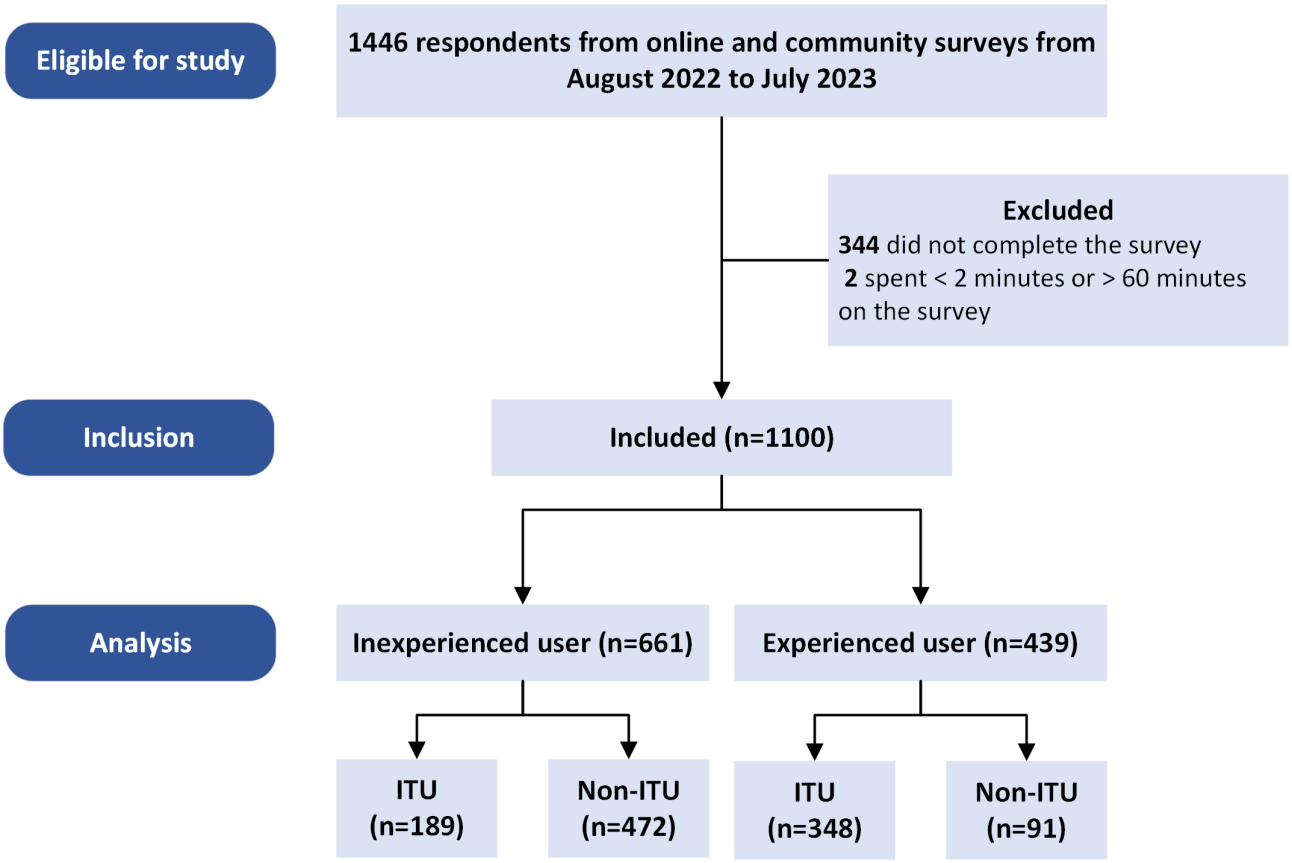
For the analysis of the primary objective, we hypothesized that the factors associated with ITU would be different for participants who had previously used mHealth and those who did not. We performed a multivariable logistic regression for 2 specified subgroups, adjusted for the confounders including age, gender, education levels, income, and living area. The predetermined characteristic variables included having any underlying disease, owning a smartphone, the number of years of experience using it, and an interaction term based on both ownership and experience. We entered each item's mHealth acceptance into the model as continuous data, using a 10-point scale. The magnitude of associations was presented using an adjusted odds ratio along with 95% CIs and *P* value. All statistical analyses were conducted using STATA (version 17.0; Stata Corp, LP). The visualization for this study was created using Microsoft Excel (Microsoft Corporation, 2021). The level of statistical significance for descriptive analysis was set at a *P* value below .05.

Results

Participant Characteristics

Of 1100 participants, the majority were female (776/1100, 62.3%), with a mean age of 62.3 (SD 8.8) years. This study flow diagram is presented in Figure 1. Most of the participants were married and resided with their families primarily in rural, suburban, and urban areas, respectively. For socioeconomic status, 86.2% (948/1100) had a low income (<10,000 Baht, US $274), and 65.9% (725/1100) had the highest education level at primary school. The common underlying diseases were hypertension and dyslipidemia. More than half reported having a vision problem, and 65.2% (n=399) wore glasses or contact lenses. Only 3.3% (4/1100) of respondents required a hearing aid, while 10.9% (120/1100) of respondents reported having hearing problems.

Figure 1. Study flow diagram. ITU: intention to use.



There were 537 participants who had ITU (48.8%) and 563 participants who did not (51.2%). When comparing the participants who had ITU mHealth apps with those who did not, the ITU group had a significantly younger average age, higher education levels, higher income, and a lower proportion of underlying diseases. Furthermore, 94.4% (507/537) and 92% (494/537) of the ITU group reported having experience using a smartphone or tablet and the internet; 91.2% (490/537) and 3.4% (18/537) had owned a smartphone and tablet, respectively. These percentages in the ITU group were significantly higher than those in the non-ITU group. The overall median (IQR) year of experience using the devices was 5 (0 - 10). When comparing the ITU to non-ITU groups, the median years of device usage experience in the ITU were significantly higher than non-ITU (median 6, IQR 3 - 10 versus median 3, IQR 0 - 9, P<.001). In addition, 66.9% of those in the ITU group had experience using mHealth apps, which was a significantly higher proportion than the non-ITU group of 17.4%. The details of participant characteristics are presented in Table 1.

Table . Participant characteristics.

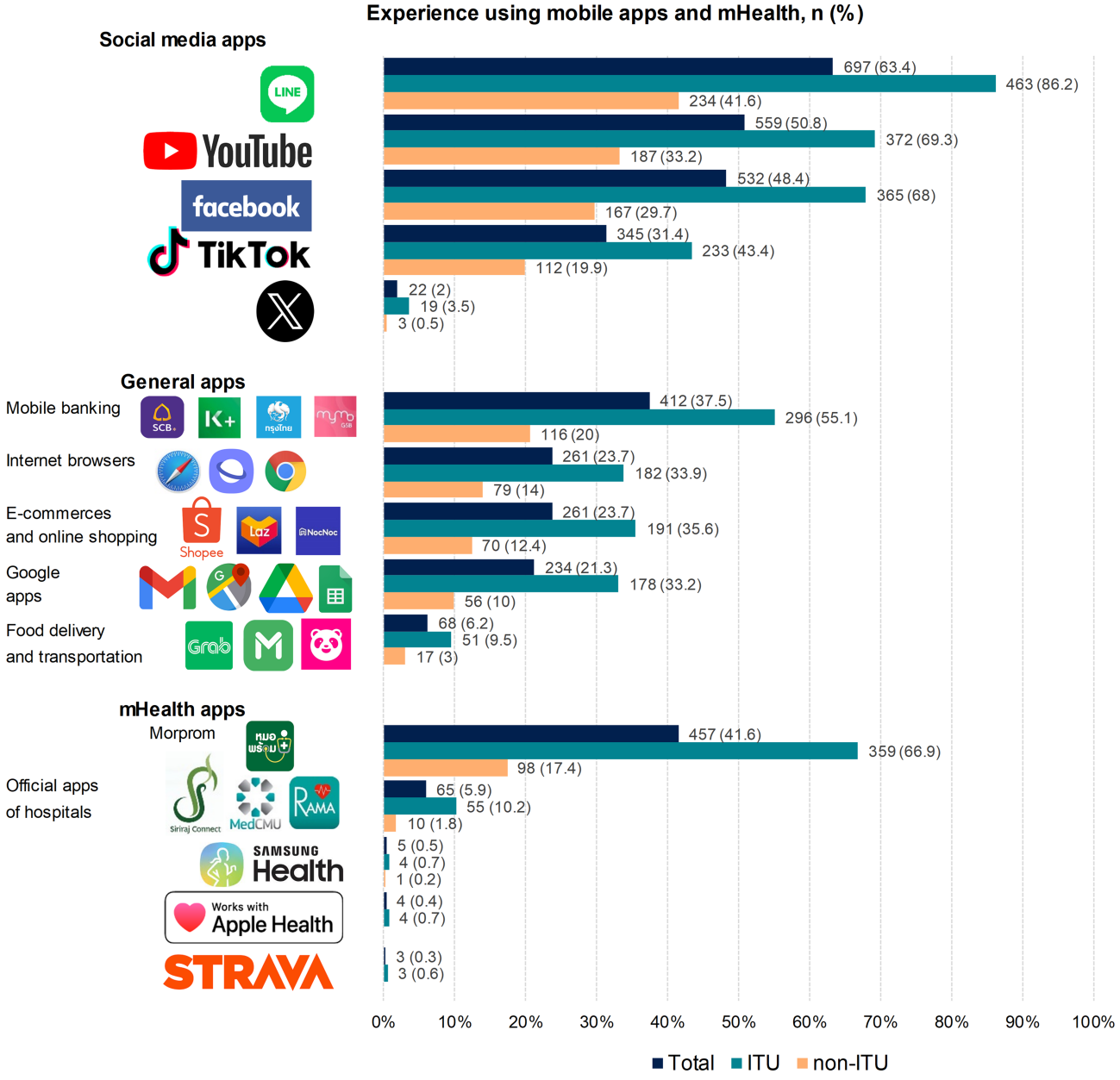
Characteristics	Total (N=1100)	Had intention to use mHealth apps		*P* value
		Yes (n=537)	No (n=563)	
Age (year), mean (SD)	62.3 (8.8)	60 (8.4)	64.4 (8.6)	<.001
Male, n (%)	324 (29.5)	140 (26.1)	184 (32.7)	.02
Marital status, n (%)				.002
Single	96 (8.7)	57 (10.6)	39 (6.9)	
Married	747 (67.9)	377 (70.2)	370 (65.7)	
Separated, divorced, or widowed	257 (23.4)	103 (19.2)	154 (27.4)	
Education levels, n (%)				<.001
No education	18 (1.6)	2 (0.4)	16 (2.8)	
Primary school	725 (65.9)	291 (54.2)	434 (77.1)	
Secondary school	97 (8.8)	65 (12.1)	32 (5.7)	
High school and vocational training	162 (14.7)	110 (20.5)	52 (9.2)	
Preuniversity	11 (1)	8 (1.5)	3 (0.5)	
Bachelor’s degree	79 (7.2)	54 (10.1)	25 (4.4)	
Master’s degree	8 (0.7)	7 (1.3)	1 (0.2)	
Income per month, n (%)				.001
<10,000 Baht (US $274)	948 (86.2)	442 (82.3)	506 (89.9)	
10,001 - 30,000 Baht (US $275 - 819)	138 (12.5)	86 (16)	52 (9.2)	
>30,001 Baht (US $820)	14 (1.3)	9 (1.7)	5 (0.9)	
Living status, n (%)				.86
Alone	108 (9.8)	50 (9.3)	58 (10.3)	
With family	988 (89.8)	485 (90.3)	503 (89.3)	
With others	4 (0.4)	2 (0.4)	2 (0.4)	
Living area, n (%)				.38
Urban	220 (20)	99 (18.4)	121 (21.5)	
Suburban	377 (34.3)	192 (35.8)	185 (32.9)	
Rural	503 (45.7)	246 (45.8)	257 (45.6)	
Had any underlying disease, n (%)	726 (66)	330 (61.5)	396 (70.3)	.002
Hypertension, n (%)	495 (45)	212 (39.5)	283 (50.3)	<.001
Dyslipidemia, n (%)	375 (34.1)	187 (34.8)	188 (33.4)	.62
Diabetes mellitus, n (%)	184 (16.7)	76 (14.2)	108 (19.2)	.03
Chronic kidney disease, n (%)	17 (1.5)	8 (1.5)	9 (1.6)	.88
Vision problems, n (%)	612 (55.6)	304 (56.6)	308 (54.7)	.53
Wore glasses or contact lens, n (%)	399 (65.2)	210 (69.1)	189 (61.4)	.045
Hearing problems, n (%)	120 (10.9)	54 (10.1)	66 (11.7)	.38
Used hearing aids, n (%)	4 (3.3)	2 (3.7)	2 (3)	.84
Number of medications, median (IQR)	1 (0 - 2)	1 (0 - 2)	1 (0 - 2)	.55
Had experience using a smartphone or tablet, n (%)	873 (79.4)	507 (94.4)	366 (65)	<.001
Had own smartphone, n (%)	843 (76.6)	490 (91.2)	353 (62.7)	<.001
Had own tablet, n (%)	20 (1.8)	18 (3.4)	2 (0.4)	<.001
Had experience using the internet, n (%)	784 (71.3)	494 (92)	290 (51.5)	<.001

Characteristics	Total (N=1100)	Had intention to use mHealth apps		
		Yes (n=537)	No (n=563)	*P* value
Had experience using mHealth apps, n (%)	439 (39.9)	348 (64.8)	91 (16.2)	<.001

Experience Using Mobile Apps and mHealth

Figure 2 illustrates participants' experience using mobile apps and mHealth. There are many types of mobile apps and mHealth including social media apps, general apps, and mHealth apps. Among the middle-aged and older adults who had previously used mobile apps and mHealth, social media apps had the highest percentages (63.4%), followed by mHealth apps (41.6%), and apps for general purposes (37.5%). The use percentages of almost all mobile apps and mHealth were statistically and significantly higher in the ITU compared to non-ITU groups (*P*<.001), except for mHealth apps with a low use percentage (Samsung Health, Apple Health, and Strava). The use percentages of the top 3 social media apps (Line, YouTube, and Facebook) were greater than other apps. For general purposes, most participants used mobile apps for online banking services, surfing the internet, and shopping. For mHealth, the Ministry of Thai Public Health app, "MorProm," had the highest use (41.6%), followed by the official apps by hospitals (5.9%), while the other mHealth apps from public providers showed a very low use percentage.

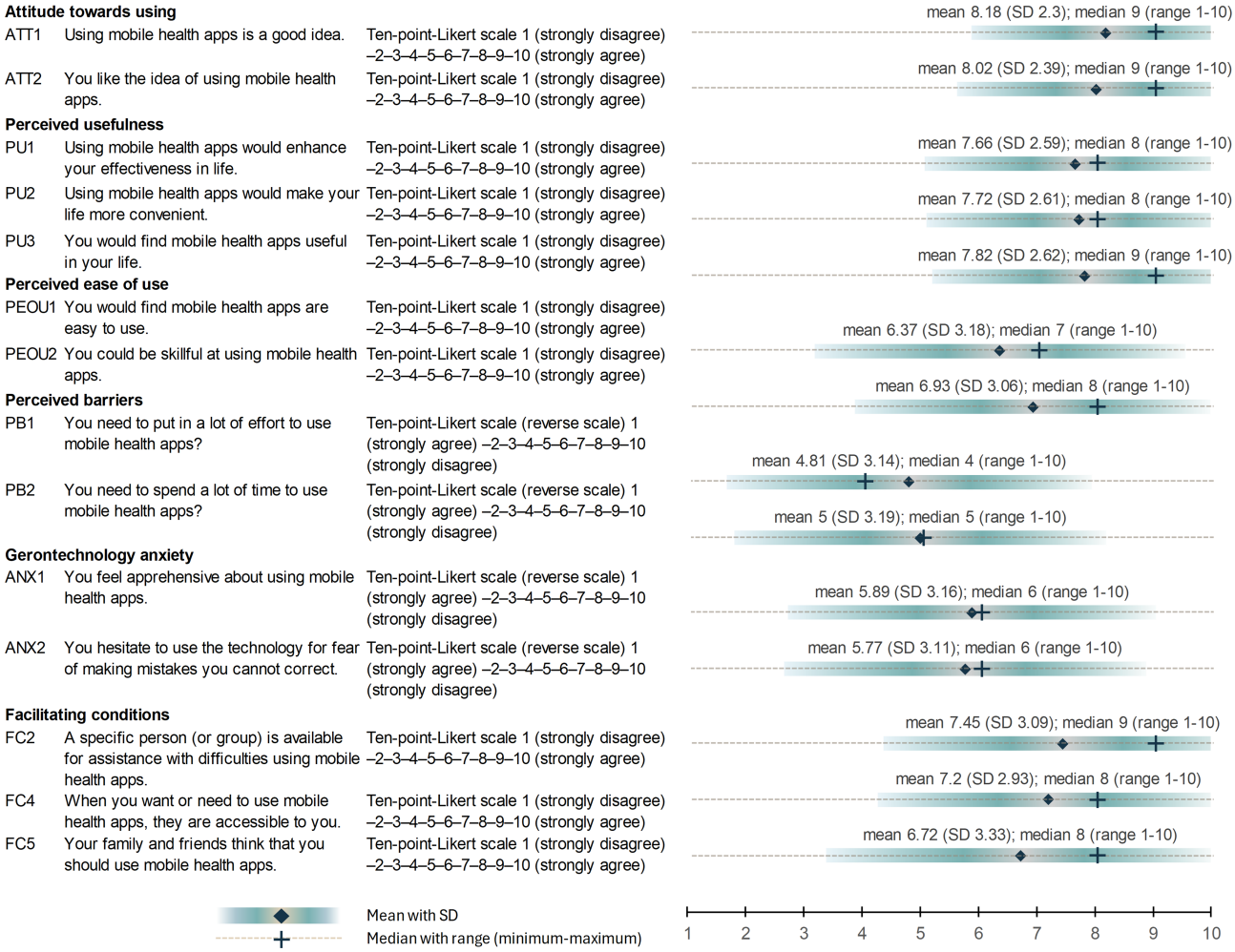
Figure 2. Participants' experience using mobile apps and mHealth. ITU: intention to use.



mHealth Acceptance

We assessed the participants' mHealth acceptance using the Thai mHealth STAM questionnaires in the domains, including attitude toward using, perceived usefulness, perceived ease of use, perceived barriers, gerontechnology anxiety, and facilitating condition. Figure 3 displays the structure of the questionnaires and the participants' responses. Overall, participants' acceptance of mHealth fell within the range of agreeable attitudes toward using, perceived usefulness of mHealth apps, and facilitating conditions. On the other hand, we observed responses in the range of neutral to low acceptance across the domains of perceived barriers, gerontechnology anxiety, and perceived ease of use.

Figure 3. Participants' responses to mHealth acceptance questionnaires. ANX: gerontechnology anxiety; ATT: attitude toward using; FC: facilitating condition; PB: perceived barriers; PEOU: perceived ease of use; PU: perceived usefulness.

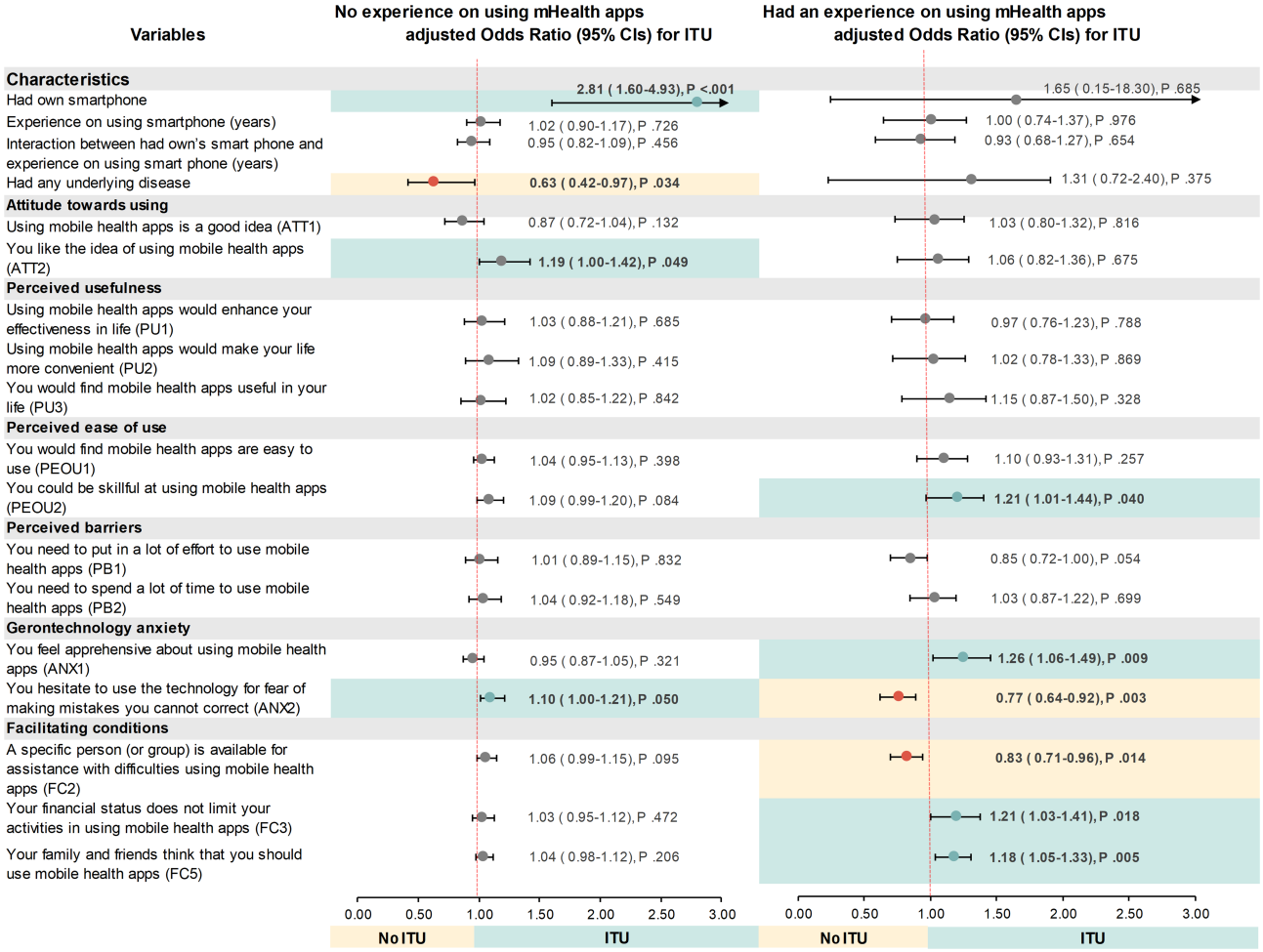


When compared between the ITU and non-ITU groups, mHealth acceptance levels assessed by the questionnaires showed a significant difference in all items, as presented in Table S1 in Multimedia Appendix 1. Participants who used an ITU mHealth app had higher mean scores in all items. The domain of perceived barriers yielded the lowest mean mHealth acceptance score in both groups.

Factors Associated With ITU mHealth Apps

In this study, we focused on exploring factors associated with ITU mHealth apps in the subgroup of participants who had experience using mHealth apps and who did not. We performed a multivariable logistic regression analysis using predetermined factors including owning a smartphone, experience using a smartphone, underlying disease, and mHeath acceptances. The analysis also adjusted for the confounders consisting of age, gender, education levels, income, and living area. Multivariable analysis results are presented in Figure 4.

Figure 4. Factors associated with intention to use mHealth apps using a multivariable logistic regression. aORs were estimated as adjusted for age, gender, education levels, income, and living area. aOR: adjusted odds ratio; ANX: gerontechnology anxiety; ATT: attitude toward using; FC: facilitating condition; ITU: intention to use; PB: perceived barriers; PEOU: perceived ease of use; PU:perceived usefulness.



For the subgroup of people who had never used mHealth apps before, having a smartphone was substantially associated with higher odds of ITU (adjusted odds ratio 2.81, 95% CI 1.60 to 4.93; *P*<.001), while having any underlying disease was significantly related to a low ITU. Higher levels of acceptance, which are characterized by a higher score of attitudes toward the idea of using mHealth and a lower fear of making mistakes that cannot be corrected when using it, were also significantly associated with a high ITU.

For the subgroup of individuals who had prior experience using mHealth apps, the pattern of associations between predetermined factors varied from the subgroup of those who did not have prior experience. There were no predetermined characteristics relating to ITU. The levels of mHealth acceptance were significantly associated with ITU in the domains, including perceived ease of use, gerontechology anxiety, and facilitating conditions. Perceived ease of use, which individuals could be skilled at using, was significantly associated with a greater odds of ITU. Regarding the level of anxiety associated with mHealth usage, less apprehensiveness about using mHealth significantly increased the odds of ITU. Conversely, a low ITU was associated with a lower fear of making mistakes when using mHealth. Having facilitated support from family and friends to use mHealth and having fewer financial constraints on using mHealth showed a significant positive association with ITU, whereas the support from others was negatively associated with ITU.

Discussion

Principal Findings

To investigate the factors associated with ITU, we found that people who had never used mHealth apps before, having a smartphone associated with a higher ITU, while having any underlying disease was significantly related to a low ITU. Higher levels of acceptance, which are characterized by a higher score of attitudes toward the idea of using mHealth and a lower fear of making mistakes, were also significantly associated with a high ITU. For the subgroup of individuals who had prior experience using mHealth apps, there were no predetermined characteristics relating to ITU. The levels of mHealth acceptance were significantly associated with ITU in the domains, including perceived ease of use, gerontechology anxiety, and facilitating conditions.

Compared to previous studies, several patient characteristics significantly influence the ITU mHealth. Older patients are less likely to use mHealth due to their unfamiliarity with technology. Despite efforts to adopt new technology, many older adult individuals still struggle with it, and the rapid pace of technological advancements leaves them unable to keep up

[16,17]. Additionally, there is a lack of promotion and support for using medical technology among older adults [18]. Socioeconomic status is another crucial factor affecting the ITU mHealth [19,20]. Patients in low-income groups not only lack support for using mHealth, but their daily routines and various burdens also reduce their intention to use these technologies [21]. This is closely related to the level of education [19]; patients with higher educational levels are more likely to understand and recognize the health benefits of mHealth, thus increasing their intention to use it [20]. Chronic diseases also play a role [19]. Patients without comorbidities and those who could effectively manage their chronic conditions tend to have better health management skills and a stronger desire to maintain good health, which positively influences their ITU mHealth [22,23].

Many factors influence the ITU mHealth among people without prior experience. One critical factor is the availability of equipment, such as mobile phones, patient testing devices, personal digital assistants, and other wireless tools. Health care professionals must ensure these facilities are accessible to patients when initiating mHealth services. Providing the necessary equipment and educating patients on its use can reduce inequity, particularly among older populations. The previous studies also showed that when health care systems provide access to such equipment, it enhances user confidence and intention to adopt mHealth, particularly among older adults who may otherwise face barriers due to digital inequality (eg, lacking smartphones or internet access) [24,25]. Empowering patients to access and use smartphones not only enhanced their ITU mHealth but could also improve self-rated, physical, and psychological health levels, as evidenced by the previous studies on the effects of smartphones and smart devices [26-28]. Underlying health conditions also impact the ITU mHealth. Patients with poor health behaviors or low motivation for self-care are less likely to adopt mHealth. The previous studies found that older adults with heart failure [29] and adults with hypertension [30] faced significant barriers to mHealth adoption, including low motivation and limited health literacy, which hindered their engagement with self-management technologies. These support our finding that underlying health conditions may reduce mHealth adoption, as individuals often lack the motivation to use health-promoting technologies. Conversely, some studies suggest that the presence of chronic health conditions can actually motivate individuals to engage with mHealth solutions. The study by Askari et al [31] found that older adults who were able to control their chronic diseases often expressed a greater ITU mHealth, as these tools can provide essential support for self-management and health monitoring. This indicated that while health conditions could pose barriers, they can also motivate the adoption of mHealth solutions, particularly when patients recognize the potential benefits for their health management. Another study also suggested that health conditions could influence older adults' readiness to engage with eHealth resources, with those experiencing chronic illnesses showing a higher willingness to use digital health tools for self-care [32]. This suggests that the context of the health condition is vital in influencing attitudes toward technology use, rather than the notion that poor health behaviors uniformly result in reduced intentions to use mHealth. Confidence in their ability to effectively use mHealth apps is crucial, as a positive attitude toward mHealth increases the likelihood of its use [33]. The positive finding on attitude toward using mHealth is also consistent with the recent systematic reviews on the impact of mHealth interventions, which highlighted that the effectiveness of these interventions is often linked to user confidence and positive attitudes toward technology [4]. To promote mHealth for people without prior experience, health care professionals should educate patients about its benefits and how it can aid in self-care. Additionally, patients must develop eHealth literacy, defined as "the ability to seek out, find, evaluate, and appraise, integrate, and apply what is gained in electronic environments toward solving a health problem." [34].

For participants with prior experience using mHealth, various factors influence their intention to continue its usage. These individuals often possess a positive attitude toward mHealth and recognize its benefits. They may also have experience in overcoming barriers, such as troubleshooting difficulties or correcting errors independently. Consequently, they perceive the benefits of using mHealth to outweigh the risks or challenges. Our finding indicated that perceived self-capability to use mHealth and lower concerns was associated with higher ITU among participants with prior mHealth experience. This finding is aligned with the study by Opoku et al [35], which emphasized that enabling resources, such as the durability and simplicity of mobile technology, significantly influence patients' perceived ease of use of mHealth interventions. This implies that when patients are informed and experienced about the user-friendly nature of mHealth apps, their confidence in using these tools increases, thereby enhancing their likelihood of adoption. In contrast to participants without prior experience using mHealth, those with experience did not avoid using mHealth out of fear of making mistakes. However, the expressed greater concerns about mHealth itself, such as its effectiveness, data security, and privacy, were associated with a low ITU. Our findings align with the studies on experienced users' attitudes toward mHealth [36,37], emphasizing the importance of addressing concerns related to security, privacy, and perceived effectiveness. Therefore, future mHealth interventions should not only focus on enhancing user confidence through health education or instruction by technical and health care providers but also on building trust in technology itself. Additionally, patients, especially older adults, value support from family or caregivers. mHealth for older adults may require assistance at various stages, including tool usage guidance, instructions, and caregiver involvement [16]. Financial considerations also play a crucial role, as increased costs associated with mHealth can affect the ITU for it. To initiate and improve mHealth usage, health care professionals need to identify patients willing to use mHealth, ensure the availability of necessary equipment, and address any financial constraints. Educating patients on the evidence and benefits of mHealth is also essential. Where possible, families and caregivers should be included in the mHealth care plan [38].

The strength of this study lies in its relatively large sample size, which included participants from rural, suburban, and urban areas. This study comprehensively explored various factors

potentially associated with mHealth use, using reliable and validated standard tools to evaluate acceptance. However, some limitations existed. The cross-sectional nature of this study limited establishing causal relationships between the ITU mHealth and associated factors. Further research designs, including longitudinal studies or randomized controlled trials of specific mHealth interventions, are necessary to overcome these limitations and provide a more comprehensive understanding of the factors influencing mHealth usage. Furthermore, the term "mHealth" was defined in a broad and general sense in the questionnaire but was not explicitly specified for participants' conditions. Defining "mHealth" in general could hinder the participants' acceptance and use of mHealth, particularly in the group of participants with underlying diseases and specific health problems. This hypothesis could explain the observed association between having any underlying disease and low ITU in the group of participants who had never used mHealth before. Accordingly, our study primarily aimed to explore mHealth usage in general for community-dwelling adults. Our findings may not generalize enough to provide specific insight into the specific use case of mHealth in a particular population. Hence, we encourage future studies to aim for an in-depth understanding of each type of mHealth and acceptance in the specific context.

Conclusion

This study identified key factors influencing the ITU mHealth and positive attitudes toward using significantly enhancing ITU in both inexperienced and experienced users. Conversely, having an underlying disease was associated with a decrease in ITU, indicating that this group of individuals may require mHealth's specific requirements for their conditions rather than the general purpose. For experienced users, acceptance levels in areas such as ease of use and gerontechnology anxiety were critical. Health care professionals should identify patients open to using mHealth, ensure access to necessary equipment, address financial barriers, and educate patients on its benefits. Future investigations should also explore specific mHealth apps to provide more targeted insights.

Acknowledgments

This work was supported by Chiang Mai University and the Faculty of Medicine Research Fund (grant 124-2565).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additonal table. The comparison of participants' responses to mHealth acceptance questionnaires.
[DOCX File, 19 KB - humanfactors_v12i1e63607_app1.docx]

References

1. Pires IM, Marques G, Garcia NM, Flórez-Revuelta F, Ponciano V, Oniani S. A research on the classification and applicability of the mobile health applications. J Pers Med 2020 Feb 27;10(1):11. [doi: 10.3390/jpm10010011] [Medline: 32120849]
2. Pires IM, Garcia NM, Pombo N, Flórez-Revuelta F, Rodríguez ND. Validation techniques for sensor data in mobile health applications. J Sens 2016;2016(1):1-9. [doi: 10.1155/2016/2839372]
3. Pires IM, Marques G, Garcia NM, et al. A review on the artificial intelligence algorithms for the recognition of activities of daily living using sensors in mobile devices. In: Singh PK, Bhargava BK, Paprzycki M, Kaushal NC, Hong WC, editors. Handbook of Wireless Sensor Networks: Issues and Challenges in Current Scenario's: Springer International Publishing; 2020:685-713.
4. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR mHealth uHealth 2018 Jan 17;6(1):e23. [doi: 10.2196/mhealth.8873] [Medline: 29343463]
5. van Elburg FRT, van de Klundert J, Nieboer AP, Askari M. The intention to use mHealth applications among Dutch older adults prior and during the COVID pandemic. Front Public Health 2023;11:1130570. [doi: 10.3389/fpubh.2023.1130570] [Medline: 37383259]
6. Buawangpong N, Pinyopornpanish K, Pliannuom S, et al. Designing telemedicine for older adults with multimorbidity: content analysis study. JMIR Aging 2024 Jan 10;7:e52031. [doi: 10.2196/52031] [Medline: 38198201]
7. Mäkinen H, Haavisto E, Havola S, Koivisto JM. User experiences of virtual reality technologies for healthcare in learning: an integrative review. Behav Inf Technol 2022 Jan 2;41(1):1-17. [doi: 10.1080/0144929X.2020.1788162]
8. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
9. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]

10. Anantanasuwong D, Theerawanviwat D, Siripanich P. Panel survey and study on health and aging, and retirement in Thailand. In: Encyclopedia of Gerontology and Population Aging: Springer; 2019, Vol. 10:978-973. [doi: 10.1007/978-3-319-69892-2_982-1]
11. Pengpid S, Peltzer K. Prevalence and associated factors of incident and persistent loneliness among middle-aged and older adults in Thailand. BMC Psychol 2023 Mar 14;11(1):70. [doi: 10.1186/s40359-023-01115-4] [Medline: 36918991]
12. Twilt M. mHealth applications are they the future or another burden for our patients? mHealth 2023;9:9. [doi: 10.21037/mhealth-23-5] [Medline: 37089269]
13. Adibi S. Mobile Health: A Technology Road Map: Springer; 2015, Vol. 5. [doi: 10.1007/978-3-319-12817-7]
14. Chen K, Chan AHS. Gerontechnology acceptance by elderly Hong Kong Chinese: a Senior Technology Acceptance Model (STAM). Ergonomics 2014;57(5):635-652. [doi: 10.1080/00140139.2014.895855] [Medline: 24655221]
15. Bujang MA, Sa'at N, Sidik T, Joo LC. Sample size guidelines for logistic regression from observational studies with large population: emphasis on the accuracy between statistics and parameters based on real life clinical data. Malays J Med Sci 2018 Jul;25(4):122-130. [doi: 10.21315/mjms2018.25.4.12] [Medline: 30914854]
16. Elbeddini A, Prabaharan T, Almasalkhi S, Tran C, Zhou Y. Barriers to conducting deprescribing in the elderly population amid the COVID-19 pandemic. Res Social Adm Pharm 2021 Jan;17(1):1942-1945. [doi: 10.1016/j.sapharm.2020.05.025] [Medline: 32499161]
17. Liu N, Yin J, Tan SSL, Ngiam KY, Teo HH. Mobile health applications for older adults: a systematic review of interface and persuasive feature design. J Am Med Inform Assoc 2021 Oct 12;28(11):2483-2501. [doi: 10.1093/jamia/ocab151] [Medline: 34472601]
18. Mace RA, Mattos MK, Vranceanu AM. Older adults can use technology: why healthcare professionals must overcome ageism in digital health. Transl Behav Med 2022 Dec 30;12(12):1102-1105. [doi: 10.1093/tbm/ibac070] [Medline: 36073770]
19. Jacob C, Sezgin E, Sanchez-Vazquez A, Ivory C. Sociotechnical factors affecting patients' adoption of mobile health tools: systematic literature review and narrative synthesis. JMIR mHealth uHealth 2022 May 5;10(5):e36284. [doi: 10.2196/36284] [Medline: 35318189]
20. Lee M, Kang D, Yoon J, et al. The difference in knowledge and attitudes of using mobile health applications between actual user and non-user among adults aged 50 and older. PLoS ONE 2020;15(10):e0241350. [doi: 10.1371/journal.pone.0241350]
21. McCool J, Dobson R, Whittaker R, Paton C. Mobile health (mHealth) in low- and middle-income countries. Annu Rev Public Health 2022 Apr 5;43:525-539. [doi: 10.1146/annurev-publhealth-052620-093850] [Medline: 34648368]
22. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181. [doi: 10.1177/2055207619882181] [Medline: 31656632]
23. Jaana M, Paré G. Comparison of mobile health technology use for self-tracking between older adults and the general adult population in Canada: cross-sectional survey. JMIR mHealth uHealth 2020 Nov 27;8(11):e24718. [doi: 10.2196/24718] [Medline: 33104517]
24. Ye Q, Deng Z, Chen Y, Liao J, Li G, Lu Y. How resource scarcity and accessibility affect patients' usage of mobile health in China: resource competition perspective. JMIR mHealth uHealth 2019 Aug 9;7(8):e13491. [doi: 10.2196/13491] [Medline: 31400104]
25. Hall CS, Fottrell E, Wilkinson S, Byass P. Assessing the impact of mHealth interventions in low- and middle-income countries--what has been shown to work? Glob Health Action 2014;7(1):25606. [doi: 10.3402/gha.v7.25606] [Medline: 25361730]
26. Li G, Jin C, Zhao B, Wu B. Smartphone use, technology affordance for healthcare and elders' life satisfaction. Front Public Health 2022;10:861897. [doi: 10.3389/fpubh.2022.861897] [Medline: 35480578]
27. Liang X, Xiong F, Xie F. The effect of smartphones on the self-rated health levels of the elderly. BMC Public Health 2022 Mar 15;22(1):508. [doi: 10.1186/s12889-022-12952-0] [Medline: 35292010]
28. Wei Y, Guo X. Impact of smart device use on objective and subjective health of older adults: findings from four provinces in China. Front Public Health 2023;11:1118207. [doi: 10.3389/fpubh.2023.1118207] [Medline: 37546303]
29. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han HR. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
30. Chen Y, Kruahong S, Elias S, et al. Racial disparities in shared decision-making and the use of mHealth technology among adults with hypertension in the 2017-2020 health information national trends survey: cross-sectional study in the United States. J Med Internet Res 2023 Sep 13;25:e47566. [doi: 10.2196/47566] [Medline: 37703088]
31. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in the Netherlands: cross-sectional study. J Med Internet Res 2020 Sep 4;22(9):e18080. [doi: 10.2196/18080] [Medline: 32624465]
32. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220. [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
33. Azhar FA, Dhillon JS. An investigation of factors influencing the intention to use mHealth apps for self-care. IJBIS 2018;29(1):59. [doi: 10.1504/IJBIS.2018.094005]

34. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9. [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
35. Opoku D, Stephani V, Quentin W. A realist review of mobile phone-based health interventions for non-communicable disease management in sub-Saharan Africa. BMC Med 2017 Feb 6;15(1):24. [doi: 10.1186/s12916-017-0782-z] [Medline: 28162090]
36. Hossain I, Lim ZZ, Ng JJL, Koh WJ, Wong PS. Public attitudes towards mobile health in Singapore: a cross-sectional study. mHealth 2018;4:41. [doi: 10.21037/mhealth.2018.09.02] [Medline: 30363776]
37. Korpershoek YJG, Vervoort SCJM, Trappenburg JCA, Schuurmans MJ. Perceptions of patients with chronic obstructive pulmonary disease and their health care providers towards using mHealth for self-management of exacerbations: a qualitative study. BMC Health Serv Res 2018 Oct 4;18(1):757. [doi: 10.1186/s12913-018-3545-4] [Medline: 30286761]
38. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mhealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]

Abbreviations

ITU: intention to use
mHealth: mobile health care
STAM: Senior Technology Acceptance Model
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by A Kushniruk; submitted 25.06.24; peer-reviewed by P Nohl-Deryk, X Zhang; revised version received 04.11.24; accepted 18.02.25; published 07.03.25.

Please cite as:
Buawangpong N, Sirikul W, Siviroj P
Factors Associated With the Intention to Use mHealth Among Thai Middle-Aged Adults and Older Adults: Cross-Sectional Study
JMIR Hum Factors 2025;12:e63607
URL: https://humanfactors.jmir.org/2025/1/e63607
doi:10.2196/63607

Original Paper

Assessing the Relationship Between the Type of Internet Use and Internet Addiction in Early and Middle Adolescents: Cross-Sectional Study From Qatar

Khansa Chemnad[1], MSc; Maryam Aziz[1], MSc; Sanaa Al- Harahsheh[2], PhD; Azza Abdelmoneium[3], PhD; Ahmed Baghdady[4], EdD; Diana Alsayed Hassan[5], PhD; Raian Ali[1], PhD

[1]College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[2]World Innovation Summit for Health, Qatar Foundation, Doha, Qatar

[3]Doha International Family Institute, Qatar Foundation, Doha, Qatar

[4]World Innovation Summit for Education, Qatar Foundation, Doha, Qatar

[5]Department of Public Health, College of Health Sciences, QU Health, Qatar University, Doha, Qatar

Corresponding Author:
Diana Alsayed Hassan, PhD
Department of Public Health
College of Health Sciences
QU Health, Qatar University
PO Box 2713
Doha
Qatar
Phone: 974 31576271
Email: dalsayed@qu.edu.qa

Abstract

Background: With the increasing prevalence of digital technology, adolescent internet addiction (IA) has become a global concern. Excessive internet use, especially among adolescents, has been linked to various negative outcomes such as poor academic performance, social isolation, and mental health issues. Conducted among adolescents of Arab origin, our study addressed the limitations of the literature, which predominantly focuses on Western, educated, industrialized, rich, and democratic populations.

Objective: This study aimed to differentiate between essential and nonessential internet use and how they relate to IA in early and middle adolescents, as well as the relationship between subjective happiness with the amount of time spent on nonessential internet use and IA.

Methods: A cross-sectional survey was conducted among 377 students from 16 schools in Qatar. The survey measured essential and nonessential internet use, subjective happiness with nonessential use, and IA symptoms using the Internet Addiction Diagnostic Questionnaire, as well as participant demographics. To explore age-specific associations, participants were categorized into early (age 11-13 years) and middle (age 14-17 years) adolescents. Factorial analysis, multiple regression, and logistic regression were used for statistical analysis.

Results: Nonessential internet use significantly predicted IA in both early (P<.001) and middle (P<.001) adolescents, with early adolescents showing a stronger association. Subjective happiness with nonessential internet use negatively predicted IA only in middle adolescents (P<.001) as greater dissatisfaction led to a higher IA risk. Essential internet use did not predict IA in either group.

Conclusions: Differentiating between essential and nonessential internet use is crucial in understanding IA. This study highlights the importance of developmental differences in shaping IA symptoms. The findings suggest that interventions aimed at addressing IA should be age specific and focus on addressing nonessential use specifically rather than considering internet use and screen time in general as a single entity. Cultural and regional factors also play a role in shaping internet use patterns and IA in the Middle East, necessitating context-specific, culturally sensitive approaches to IA prevention.

(JMIR Hum Factors 2025;12:e62955) doi:10.2196/62955

KEYWORDS

internet addiction; internet use; early adolescence; middle adolescence; mobile phone

Introduction

Background

Adolescent internet addiction (IA) has become a growing concern in today's society as the use of technology and the internet has become increasingly prevalent in the lives of young people. According to a study conducted by the Pew Research Center, since 2014 to 2015, there has been a modest increase in the proportion of adolescents who said that they used the internet daily or more often [1]. In a follow-up study in 2022, a total of 97% of adolescents said that they used the internet daily compared to 92% of adolescents who said the same in 2014 to 2015. It is understandable that many adolescents struggle with IA given the widespread use of the internet. Studies have shown that approximately 10% of adolescents are at risk of IA, with even higher rates reported in countries such as Hong Kong, China, and South Korea [2-5]. For example, a survey conducted in South Korea found that nearly 30% of junior high school students were either addicted to the internet or at a risk of IA [6]. While some internet use can be beneficial for educational and social purposes, excessive use can lead to negative consequences such as decreased academic performance, social isolation, and even mental health issues [7-9]. Therefore, it is important to understand the prevalence of IA among adolescents and its contributing factors.

IA is defined as the inability to control internet use, leading to disruptions in everyday functioning and symptoms of withdrawal and tolerance [10]. Although not officially recognized as a disorder in the most recent edition of the *Diagnostic and Statistical Manual of Mental Disorders*, IA has received considerable attention in scientific literature and is regarded as a growing concern [11]. Opinions in the literature differ, with some researchers suggesting distinct diagnostic criteria and subtypes of IA, such as video game and online gambling addiction [12]. However, there is still a lack of consensus regarding the internet's addictive nature and IA diagnostic criteria [13]. While there is still debate regarding the exact nature and criteria for IA, it is clear that excessive internet use can have negative consequences, particularly for adolescents. Understanding the prevalence and factors contributing to IA is important for addressing this growing concern, promoting healthy internet use, and developing appropriate interventions for those who may be affected.

Most psychological research studies predominantly rely on samples from Western, educated, industrialized, rich, and democratic (WEIRD) populations [14]. However, when examining the effects of excessive internet use on adolescent mental health, it is crucial to consider that the experiences of adolescents from non-WEIRD populations may vary owing to many factors, such as different cultural norms and access to technology [15]. This emphasizes the importance of including more diverse samples in research to gain a comprehensive understanding of the global impact of technology on mental health. Cultural context significantly shapes adolescence, with the experiences of young individuals varying widely across different cultures [16]. To address this gap, our study focused on adolescent IA in the Middle East, where limited research on IA has been conducted. In doing so, we aimed to shed light on the specific dynamics of IA in this region and contribute to a more inclusive and comprehensive understanding of this phenomenon.

Qatar has one of the highest internet penetration rates in the world at 99% [17]. The country's internet users have grown significantly from 69% in 2012 to 100% in 2022 [18]. This increase raises the possibility of an increase in IA prevalence rates, as does the spread of mobile device use and the constant release of new technology updates. Furthermore, adolescents may unintentionally be encouraged to spend more time inside streaming online content, playing online games, and spending time on-screen due to Qatar's humid and hot weather, which lasts for more than half the year.

In addition to Qatar's unique technological landscape and climate, broader Middle Eastern cultural factors play a crucial role in shaping adolescent internet use patterns. The region's complex environment, which is characterized by its unique family dynamics, societal expectations, and cultural values, may contribute to the elevated prevalence of IA among adolescents [19,20]. In various Middle Eastern societies, conservative norms that prioritize modesty and restraint frequently conflict with the liberties that online platforms provide. As adolescents navigate their identities and seek communities that are consistent with their evolving values, this cultural tension can motivate them to engage with the internet differently. In addition, it is also argued that the limited opportunities for social interaction outside the home, along with societal pressures to conform, may encourage adolescents to seek connection and validation online, potentially nurturing addictive behaviors [19]. Culturally informed research can inform tailored intervention strategies that address the specific needs of adolescents in the Middle East. For example, studies suggest that effective interventions should involve families and schools, recognizing the significant role that these environments play in shaping adolescents' internet behaviors [21,22].

Excessive screen time is often identified as a prominent indicator of IA, contributing to an unhealthy relationship with the internet and development of negative habits [23,24]. In numerous studies examining adolescent IA, the amount of time spent on the web is commonly regarded as a key factor in determining the presence of IA [25,26]. The COVID-19 pandemic has further heightened the reliance on technology for remote learning and socializing among adolescents [27]. The term *screen time* is frequently used without sufficient attention being paid to the specific activities undertaken during that period. However, this lack of specificity has several methodological limitations. To address these limitations, this study aimed to differentiate between the time spent on essential and nonessential internet use activities, providing a more nuanced understanding of IA in adolescents. However, IA is not just a matter of excessive

screen time; it has also been linked to mental health issues such as loneliness, low self-esteem, sleep problems, depression, social phobia, and anxiety [28,29]. Adolescents who struggle with these conditions may turn to the internet as a coping mechanism, further exacerbating their addiction [30]. Ultimately, IA can have a significant negative impact on adolescents' lives, from decreased academic performance to strained relationships with loved ones [31].

Adolescence is a crucial period for identity exploration and formation during which individuals may encounter situations in which their emerging beliefs and values conflict with societal expectations or their own behaviors, which may lead to cognitive dissonance [32]. This aspect of adolescent development is relevant to the exploration of subjective feelings regarding digital technology use and its association with IA. Cognitive dissonance theory offers an understanding of this relationship by examining potential conflicts arising from adolescents' subjective feelings regarding their digital technology use and addictive behavior. According to the cognitive dissonance theory, individuals experience psychological discomfort when they hold contradictory beliefs, attitudes, or behaviors [33]. In the context of digital technology use and IA, cognitive dissonance may arise when adolescents recognize the negative consequences of excessive digital technology use but continue to engage in it. The literature suggests that subjective positive feelings about one's own life conditions are negatively correlated with IA [34]. It is reasonable to hypothesize that similar correlations may exist between subjective feelings regarding digital technology use and IA. However, there is a lack of specific studies exploring this association. Given this gap in the literature, this study aimed to address this by examining the association between subjective feelings about digital technology use and IA.

Adolescents are growing up in an era characterized by rapid technological advancements, and the use of smartphones has become a common part of their daily lives. This pervasive technology has greatly influenced their interactions with the internet, particularly social networking sites [35]. Consequently, this has had a profound impact on various aspects of their lives, including social dynamics, cognitive development, and emotional well-being. The literature indicates notable distinctions between early and middle adolescents in terms of cognitive and psychological development [36]. It is critical to distinguish between early and middle adolescence when investigating adolescent IA owing to the distinct developmental traits, experiences, and coping strategies that arise throughout these separate phases of adolescence [37]. Early adolescence is a critical period defined by the start of puberty and the transition from childhood to adolescence. This stage is distinguished by rapid physical, cognitive, and social changes, as well as identity exploration and self-concept construction [38,39]. Furthermore, early adolescence is a critical time for the initiation and development of mental health issues [40]. Middle adolescence is characterized by managing the complications of moving into young adulthood [41]. Individuals become more self-sufficient, engage in complex social connections, and gain feelings of autonomy [42]. If these aspects of development are not addressed properly during adolescence, consequences may have a lasting impact in their adulthood. Researchers must examine the diverse vulnerabilities, risk factors, and patterns of internet use among early and middle adolescents while studying IA during these different phases. Early adolescents may be more vulnerable to IA because they are constantly trying new things, including online platforms, to define their identities and connect with friends [39]. Middle adolescence, on the other hand, is characterized by the emergence of more stable patterns of behavior, including internet use, as adolescents approach young adulthood and deal with increasing academic and social demands [43]. Furthermore, research has shown that the characteristics that influence IA change between early and middle adolescence [21]. For example, parental supervision and advice may have a greater effect among early adolescents, but peer influence and social support may become more significant in middle adolescence [44,45]. Furthermore, in middle adolescence, the onset of academic commitment and the need for self-regulation might alter internet use patterns and addiction risks. IA has become a growing concern among adolescents, with symptoms presenting both in relation to oneself and in interactions with others. However, the cognitive development and behavioral patterns of early and middle adolescents differ significantly, suggesting the need for a critical examination of IA symptoms according to age. Although previous research has highlighted the multifaceted nature of IA symptoms, little attention has been paid to the potential differential manifestation of symptoms based on the age of adolescents. Understanding how IA symptoms manifest differently in early and middle adolescence is crucial for tailoring prevention and intervention efforts. Understanding the differences between early and middle adolescence is critical when examining IA. Our study aimed to address the existing knowledge gap by investigating aspects that have not been thoroughly explored in the field of adolescent IA.

Adolescent technology use can be categorized into distinct types depending on type of use: essential and nonessential [46,47]. Activities that are essential for everyday life, such as educational pursuits or personal improvement–related activities, are included in essential use. On the other hand, nonessential use includes leisure activities such as gaming, social media engagement for nonspecific developmental purpose, and other forms of entertainment [46]. Although researchers have investigated the correlations between excessive internet use and IA, there is a noticeable absence of studies that explicitly distinguish between essential and nonessential use patterns. We suggest that the use of the internet for essential activities related to education and personal development can improve efficiency and time management and may not be linked to IA. In contrast, the excessive use of the internet for nonessential activities may help change mood and increase bonding but may also result in diminished productivity and potential adverse consequences [48]. Given this distinction, our study endeavored to examine the distinct effects of essential versus nonessential internet use on adolescents, thereby contributing to resolving a critical gap in the current literature and offering valuable insights for comprehending healthy internet use patterns.

Objectives

While extensive research has been conducted on IA among adolescents, several significant gaps remain. First, there is a lack of studies focusing on non-WEIRD populations, particularly in the Middle East, where unique cultural factors may influence IA. Second, the potential significance of the distinction between essential and nonessential internet use in relation to IA has not been adequately investigated. Third, there is a scarcity of research on the correlation between subjective emotions related to the use of digital technology and IA. Finally, while the developmental differences between early and middle adolescence are well documented, there are few studies that have investigated the potential differences in the manifestation of IA symptoms and their relationship to technology use across these age groups. We aimed to answer the following three research questions (RQs) through this study:

1. Do essential and nonessential types of internet use predict IA in early and middle adolescents? (RQ 1)
2. Can subjective happiness with nonessential internet use predict adolescent IA? (RQ 2)
3. Does the impact of time spent on essential and nonessential internet use, as well as subjective feelings regarding nonessential internet use, on IA differ between the early and middle adolescence stages? (RQ 3)

Methods

Participants

This study involved school students residing in Qatar, and their participation was secured through web-based surveys administered via SurveyMonkey (SurveyMonkey Inc). The survey invitation was sent out to multiple schools in Qatar through an open call using the mailing list of head teachers and educators. One of the authors works at an education policy institute with a network of schools that attend events and participate in research activities. This facilitated outreach. A total of 16 schools accepted the study invitation, comprising an equal number of public (n=8, 50%) and private (n=8, 50%) schools. Public schools in Qatar are government funded and managed by the government and primarily enroll Qatari citizens as they provide free education with Arabic as the language of instruction. These schools are also gender segregated, meaning that male and female students attend separate institutions. Private schools, on the other hand, are funded by tuition fees and managed by private organizations and predominantly enroll expatriate students. These schools follow a variety of government-reviewed international curricula and are typically not gender segregated. Data were collected between March 2022 and May 2022.

Before administering the survey, students were provided with an induction session by their head teacher, who explained the distinction between essential and nonessential internet use. Essential use was defined as internet activities connected to study, academic work, and personal development, whereas nonessential use was defined as activities that do not directly contribute to these aims. This briefing was designed to establish a common understanding among participants, guaranteeing that their answers precisely reflected these definitions. The authors acknowledge the difficulty in standardizing what adolescents consider essential versus nonessential. Therefore, our study and results should be interpreted as reflecting the subjective views and self-reports of the adolescent participants. This interpretation follows an induction on general understanding, which aimed to reduce diversity in perceptions.

We used the formula by Green [49] to calculate the optimal sample size for our research. The formula indicates that a linear regression analysis requires a minimum sample size of 50 plus 8 times the number of independent variables p ($50 + 8 \times p$). This indicates that a minimum sample size of 74 participants was suitable for our study. In addition, to ensure correlation stability [50], our aim was to have a sample size of >250 individuals. This sample size is consistent with recommendations for maintaining statistical power in studies with multiple predictors [49,50]. Furthermore, the male sample was deemed too small to segregate the analysis by gender.

After excluding participants who did not meet the inclusion criteria, such as those who provided incomplete responses, 64.3% (377/586) of the initial students were included in the analysis, which fulfilled the criteria of a 5% margin of error at the 95% CI. Of these 377 students, 86 (22.8%) were male and 291 (77.2%) were female. The gender distribution in our sample reflects the timing of data collection, which coincided with final exams, as well as the demographic characteristics of the responding schools. Specifically, 12% (2/16) of the schools that responded to the survey were predominantly female-only public schools, which significantly contributed to the observed gender imbalance. It is important to note that gender was not included as a variable in our analysis as the study's primary focus was on essential and nonessential internet use. On the basis of the estimates of the sample size and the need for adequate statistical power in our regression analysis, it was not possible to divide the sample by gender for separate analyses.

The average age of the participating adolescents was 13.19 (SD 1.24; range 11-17) years. The participants in the study were requested to explicitly provide their ages. Participants were categorized into 2 groups: early adolescents (aged 11-13 years) and middle adolescents (aged 14-17 years). The classification of adolescents into these developmental phases was determined based on the existing literature [51].

Ethical Considerations

This study was approved by the institutional review board of Hamad Bin Khalifa University (QBRI-IRB 2021-05-094). School permissions were obtained to distribute the survey. The parents and adolescents were informed beforehand of the study, and their written informed consent and assent were collected. Participants were informed about the study's purpose, and their involvement was voluntary, allowing them to skip questions or withdraw from the survey if desired. During data collection, no identifiable information was collected. The researchers ensured that all deidentified data were stored on a password-protected laptop that only the primary investigators of the study had access to. Participants were not compensated for taking part in this study.

Measures

Overview

The survey included questions regarding participants' demographics, digital technology use, subjective happiness with nonessential internet use, and IA. The questionnaire was distributed in both English and Arabic to accommodate language preferences, and the accuracy of the Arabic translation was ensured using the back translation method [52] (see Multimedia Appendix 1 for both versions of the survey). Our study included a comprehensive and multistep procedure for validating the survey questions to establish its reliability and suitability for the adolescent population. We began by conducting think-aloud protocols with a subset of 3 adolescents, asking them to verbalize their thought processes while answering the survey questions [53]. This approach helped us identify potential ambiguities or misunderstandings, leading to revisions in translation that ensured clarity and suitability for the target age group yet without departing from the original meaning of the questions in the English version. In total, 3 authors with extensive expertise dealing with adolescent groups in the Arab area offered expert advice. Furthermore, the questions in the survey were culturally appropriate, did not use any jargon, and were free of unclear language.

Internet Addiction Diagnostic Questionnaire

The Internet Addiction Diagnostic Questionnaire (IADQ) consists of 8 items, each representing a symptom used to identify IA [10]. The questionnaire uses a binary response format, with participants indicating *no* or *yes* to each symptom. The total score on the IADQ ranges from 0 to 8 and is obtained by summing the values assigned to each of the 8 binary questions. The symptoms assessed by the IADQ are *preoccupation* (question 1), *tolerance* (question 2), *unsuccessful efforts to limit or stop Internet usage* (question 3), *withdrawal* (question 4), *loss of control of time spent on the Internet* (question 5), *risk/lose relationships or opportunities* (question 6), *lies to conceal the extent of involvement* (question 7), and *dysfunctional coping* (question 8) [54]. Participants who answered *yes* to ≥5 symptoms were categorized as addicted to the internet (dependent internet users), whereas the others were categorized as nondependent internet users. Previous studies on the IADQ have reported Cronbach α values ranging from 0.60 to 0.72 [55]. The reliability of the IADQ in this study was deemed acceptable, with a Cronbach α value of 0.66 [56] and a McDonald ω value of 0.67. In addition, the research conducted by Lozano-Blasco et al [57] found that using lesser answer categories led to decreased Cronbach α values. The dichotomous character of the questions in the IADQ, which means that they only allow for replies of *yes* or *no*, leads to lower Cronbach α values that are within the acceptable range for dependability. To establish content validity [58], the initial item pool was reviewed by a panel of 3 experts who has extensive experience working with adolescents in the Arab region. This process resulted a questionnaire that comprehensively covered the construct of interest.

Essential and Nonessential Internet Use

Adolescents were presented with a set of questions regarding their use of digital technology for both essential and nonessential purposes on both weekends and weekdays. The weekend is officially observed on Fridays and Saturdays in Qatar. By focusing on weekdays and weekends separately, we aimed to capture any behavioral variations influenced by the shift from weekend to workweek and vice versa. Participants were asked the following questions to encompass four different situations: (1) How many hours do you use digital technology for study purposes daily on weekdays (Sunday-Thursday)? (2) How many hours do you use digital technology for nonessential reasons daily on weekdays (Sunday-Thursday)? (3) How many hours do you use digital technology for study purposes daily on weekends (Friday and Saturday)? (4) How many hours do you use digital technology for nonessential reasons daily on weekends (Friday and Saturday)?

Essential internet use on weekdays and weekends was combined, and their average was used to compute essential internet use time. Nonessential internet use on weekdays and weekends was combined, and their average was taken to compute nonessential internet use time.

Subjective Happiness With Nonessential Internet Use Time

Adolescents were asked to express their feelings regarding the extent of their nonessential use of digital technology. Their responses were collected using a five-point Likert scale: (1) "I am happy with it," (2) "I am somewhat happy with it," (3) "Neither happy nor unhappy with it," (4) "I am somewhat unhappy with it," and (5) "I am unhappy with it."

Statistical Analysis

Multiple linear regression analyses were conducted separately for early and middle adolescents to examine the relationship between nonessential internet use, essential internet use, and subjective happiness with nonessential internet use time and their IA status. The significance level for the statistical tests was set at $P<.05$. All statistical analyses were carried out using JASP (version 0.17.1) [59].

Exploratory factor analysis with oblique (promax) rotation was performed [60]. The determination of the number of significant components to retain for rotation was based on two criteria: (1) the scree plot indicating the number of extracted factors and (2) ensuring that the factor solution allowed for a coherent interpretation of the results.

Results

Descriptive Statistics

The descriptive statistics of the participants are presented in Table 1.

Table 1. Descriptive statistics—internet use patterns and internet addiction (IA).

Variable	Early adolescents (n=242)	Middle adolescents (n=135)
Essential internet use time (h), mean (SD)	2.22 (1.76)	2.96 (2.12)
Nonessential internet use time (h), mean (SD)	4.45 (2.74)	4.52 (2.35)
Subjective happiness with nonessential internet use time, n (%)		
"I am happy with it" (rating of 1)	90 (37.2)	44 (32.6)
"I am somewhat happy with it" (rating of 2)	77 (31.8)	40 (29.6)
"Neither happy nor unhappy with it" (rating of 3)	46 (19)	35 (25.9)
"I am somewhat unhappy with it" (rating of 4)	17 (7)	11 (8.1)
"I am unhappy with it" (rating of 5)	12 (5)	5 (3.7)
IA		
Total IA score (0-8), mean (SD)	3.26 (1.98)	3.47 (2.17)
IA prevalence, n (%)		
Addicted internet users	63 (26)	50 (37)
Nonaddicted internet users	179 (74)	85 (63)

Exploratory Factor Analysis of IA Symptoms

To identify underlying dimensions within the IA symptoms and reduce the data into meaningful factor scores, an exploratory factor analysis was conducted on the dataset. The Kaiser-Meyer-Olkin (KMO) test, a measure of sampling adequacy, was used in this study to evaluate multicollinearity in the data to determine the feasibility of conducting a factor analysis. The overall KMO measure was 0.76, denoting "middling" adequacy. Individual KMO measures were all of >0.68, ranging from the "mediocre" to the "marvelous" categories according to the classification system by Kaiser and Rice [61]. The Bartlett test of sphericity was statistically significant (P<.001), indicating that the data were likely factorizable.

The analysis extracted 2 components, which accounted for 26% of the variance. Our scree plot was compatible with the 2-factor model. We interpreted the 2 factors as representing internal IA symptoms (factor 1; 18% of the total variance)—those symptoms primarily related to the individual's self, such as preoccupation with internet use—and external IA symptoms (factor 2; 9% of the total variance)—those symptoms primarily related to interaction with others. Factor loadings for factor 1 and factor 2 ranged from 0.31 to 0.55 and from 0.33 to 0.78, respectively. Factor scores were derived by adding the scores of individual items within each empirical domain (sum score) and dividing the sum scores by the total number of items (mean item score). The component loadings and uniqueness of the rotated solutions are presented in Table 2.

Table 2. Factor analysis of internet addiction symptoms.

Item	Factor 1 loadings	Factor 2 loadings	Uniqueness
"Preoccupation" (question 1)	0.49	—[a]	0.79
"Tolerance" (question 2)	0.31	—	0.91
"Unsuccessful efforts to limit or stop Internet usage" (question 3)	0.50	—	0.76
"Withdrawal" (question 4)	0.55	—	0.62
"Loss of control of time spent on the Internet" (question 5)	0.50	—	0.77
"Risk/lose relationships or opportunities" (question 6)	—	0.78	0.48
"Lies to conceal the extent of involvement" (question 7)	—	0.33	0.75
"Dysfunctional coping" (question 8)	0.45	—	0.79

[a]The factor does not comprise this item.

Early Adolescents

Predictors of Total IA Scores

Overview

Multiple regression analysis was used to determine factors that predicted early adolescents' total IA scores and internal IA symptoms. No outliers were observed in the data. The Pearson correlation was also used to analyze the associations between total IA, which was the dependent variable, and the independent variables of subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time.

Correlation Analysis of Variables

The correlations between the variables are presented in Table 3.

Table 3. Pearson correlation table between the variables in early adolescents.

	Total IA[a]	Internal IA symptoms	External IA symptoms	Subjective happiness with nonessential internet use time	Essential internet use time	Nonessential internet use time
Total IA	1	—[b]	—	—	—	—
Internal IA symptoms	0.95[c]	1	—	—	—	—
External IA symptoms	0.58[c]	0.30[c]	1	—	—	—
Subjective happiness with nonessential internet use time	0.04	0.03	0.03	1	—	—
Essential internet use time	0.07	0.04	0.10	0.04	1	—
Nonessential internet use time	0.41[c]	0.41[c]	0.18[d]	0.07	0.12	1

[a]IA: internet addiction.

[b]Not applicable.

[c]$P<.001$.

[d]$P<.01$.

Regression Analysis of Total IA Scores

Table 4 shows the results of the multiple regression that was run to predict the total IA score from the subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time among early adolescents. The linearity of the sample was assessed by examining the residuals versus predicted plots, whereas residuals versus dependent plots were used to evaluate homoscedasticity in the data. All the assumptions of linearity, normality, homoscedasticity, and multicollinearity were satisfied. Residuals were independent, as assessed via a Durbin-Watson statistic of 1.93. There was no evidence of multicollinearity, as assessed via tolerance values of >0.1. The assumption of normality was met, as assessed using a *Q*-*Q* plot. The multiple regression model significantly predicted the total IA score ($F_{3, 238}$=15.72; $P<.001$; R^2=0.17; adjusted R^2=0.15). Within the model, nonessential internet use time (β=.40; $P<.001$) was the only significant predictor of early adolescents' total IA scores.

Table 4. Multiple linear regression analysis predicting internet addiction in early adolescents[a].

Predictor	Standardized	*t* test (*df*)	*P* value
Constant	—[b]	5.73 (3)	<.001
Subjective happiness with nonessential internet use time	.00923	0.16 (3)	.88
Essential internet use time	.02	0.36 (3)	.72
Nonessential internet use time	.40	6.74 (3)	<.001

[a]R^2=0.17; adjusted R^2=0.15; $F_{3, 238}$=15.72.

[b]Not applicable.

Predictors of Internal and External IA Symptoms

Multiple regression analysis was used to determine factors that predicted internal IA symptoms with the independent variables of subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time, and an ordinal logistic regression model was used to predict external IA symptoms. Table 5 shows the results of the multiple regression that was run to predict the internal IA symptom score from the subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time. The linearity of the sample was assessed by examining the residuals versus predicted plots, whereas residuals versus dependent plots were used to evaluate homoscedasticity in the data. All the assumptions of linearity, normality, homoscedasticity, and multicollinearity were satisfied. Residuals were independent, as assessed via a Durbin-Watson statistic of 1.98. There was no evidence of multicollinearity, as assessed via tolerance values of >0.1. The assumption of normality was met, as assessed using a *Q*-*Q* plot. The multiple regression model significantly predicted the internal IA symptom score ($F_{3, 238}$=15.95; $P<.001$; R^2=0.17; adjusted R^2=0.16). Within the model, nonessential internet use time (β=.41; $P<.001$) was the only significant predictor of early adolescents' internal IA symptoms.

Table 5. Multiple regression analysis predicting internal internet addiction symptoms in early adolescents[a].

Predictor	Standardized	*t* test (*df*)	*P* value
Constant	—[b]	6.18 (3)	<.001
Subjective happiness with nonessential internet use time	.00318	0.05 (3)	.96
Essential internet use time	−.00431	−0.07 (3)	.94
Nonessential internet use time	.41	6.86 (3)	<.001

[a]R^2=0.17; adjusted R^2=0.16; $F_{3, 238}$=15.95.

[b]Not applicable.

An ordinal logistic regression analysis was conducted to examine the relationship between the 3 predictor variables (ie, subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time) and the outcome variable (external IA symptoms; Table 6). The variables included in the ordinal logistic regression analysis met the assumptions of the statistical model—the dependent variable (ie, external IA symptoms) was treated as an ordinal variable, and the independent variables (ie, subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time) were treated as continuous variables. The assumption of no multicollinearity was assessed by examining the variance inflation factors (VIFs), and the result (VIFs<1.1) indicated no multicollinearity. The results revealed that nonessential internet use time was associated with increased external IA symptoms (odds ratio [OR] 1.13, 95% CI 1.02-1.24; *P*=.02). Meanwhile, essential internet use time was not associated with any change in external IA symptoms (OR 1.11, 95% CI 0.96-1.28; *P*=.17). Similarly, subjective happiness with nonessential internet use time was not associated with any change in external IA symptoms (OR 0.99, 95% CI 0.78-1.25; *P*=.94) in early adolescents.

Table 6. Ordinal logistic regression predicting external internet addiction symptoms in early adolescents.

Predictor	OR[a] (95% CI; SE)	*P* value
Subjective happiness with nonessential internet use time	0.99 (0.78-1.25; −0.01)	.94
Essential internet use time	1.11 (0.96-1.28; 0.10)	.17
Nonessential internet use time	1.13 (1.02-1.24; 0.12)	.02

[a]OR: odds ratio.

Binary logistic regression was conducted to examine the relationship between subjective happiness with nonessential internet use time and the likelihood of being a dependent internet user (addicted) compared to a nondependent internet user (unaddicted) in early adolescents. The reference category, representing nondependent internet users, was coded as class 1. The OR of 0.93 (95% CI 0.73-1.20) indicated the effect size. However, the results did not reveal a significant association between subjective happiness with nonessential internet use time and IA (*P*=.58).

Middle Adolescents

Predictors of Total IA Scores

Overview

Multiple regression analysis was used to determine the factors that predicted middle adolescents' total IA scores and internal IA symptoms. No outliers were observed in the data. The Pearson correlation was also used to analyze the associations between total IA, which was the dependent variable, and the independent variables of subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time.

Correlation Analysis of Variables

The correlations between the variables are presented in Table 7.

Table 7. Pearson correlation table between the variables—middle adolescents.

	Total IA[a]	Internal IA symptoms	External IA symptoms	Subjective happiness with nonessential internet use time	Essential internet use time	Nonessential internet use time
Total IA	1	—[b]	—	—	—	—
Internal IA symptoms	0.95[c]	1	—	—	—	—
External IA symptoms	0.66[c]	0.39[c]	1	—	—	—
Subjective happiness with nonessential internet use time	0.34[c]	0.28[d]	0.32[c]	1	—	—
Essential internet use time	0.21[e]	0.15	0.26[d]	0.18[e]	1	—
Nonessential internet use time	0.35[c]	0.33[c]	0.24[d]	0.13	0.05	1

[a]IA: internet addiction.

[b]Not applicable.

[c]P<.001.

[d]P<.01.

[e]P<.05.

Regression Analysis of Total IA Scores

Table 8 shows the results of the multiple regression that was run to predict middle adolescents' total IA score from the subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time. All the assumptions of linearity, normality, homoscedasticity, and multicollinearity were satisfied. The linearity of the sample was assessed by examining the residuals versus predicted plots, whereas residuals versus dependent plots were used to evaluate homoscedasticity in the data. Residuals were independent, as assessed via a Durbin-Watson statistic of 1.83. There was no evidence of multicollinearity, as assessed via tolerance values of >0.1. The assumption of normality was met, as assessed using a *Q-Q* plot. The multiple regression model significantly predicted the total IA score ($F_{3,\ 131}$=12.96; P<.001; R^2=0.23; adjusted R^2=0.21). Within the model, nonessential internet use time (β=.31; P<.001) and subjective happiness with nonessential internet use time (β=.27; P<.001) were significant predictors of middle adolescents' total IA score.

Table 8. Multiple linear regression analysis predicting internet addiction in middle adolescents[a].

Predictor	Standardized	*t* test (*df*)	*P* value
Constant	—[b]	1.16 (3)	.25
Subjective happiness with nonessential internet use time	.27	3.44 (3)	<.001
Essential internet use time	.15	1.90 (3)	.06
Nonessential internet use time	.31	3.94 (3)	<.001

[a]R^2=0.23; adjusted R^2=0.21; $F_{3,\ 131}$=12.96.

[b]Not applicable.

Predictors of Internal and External IA Symptoms

Multiple regression analysis was used to determine factors that predicted internal IA symptoms with the independent variables of subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time, and ordinal logistic regression was used to predict external IA symptoms. Table 9 presents the results of the multiple regression. The linearity of the sample was assessed by examining the residuals versus predicted plots, whereas residuals versus dependent plots were used to evaluate homoscedasticity in the data. All the assumptions of linearity, normality, homoscedasticity, and multicollinearity were satisfied. Residuals were independent, as assessed via a Durbin-Watson statistic of 1.82. There was no evidence of multicollinearity, as assessed via tolerance values of >0.1. The assumption of normality was met, as assessed using a *Q-Q* plot. The multiple regression model significantly predicted the internal IA symptom score ($F_{3,\ 131}$=9.15; P<.001; R^2=0.17; adjusted R^2=0.15). Within the model, nonessential internet use time (β=.29; P<.001) and subjective happiness with nonessential internet use time (β=.22; P=.007) were significant predictors of middle adolescents' internal IA symptoms.

Table 9. Multiple regression analysis predicting internal internet addiction symptoms in middle adolescents[a].

Predictor	Standardized	*t* test (*df*)	*P* value
Constant	—[b]	2.10 (3)	.04
Subjective happiness with nonessential internet use time	.22	2.76 (3)	.007
Essential internet use time	.10	1.20 (3)	.23
Nonessential internet use time	.29	3.65 (3)	<.001

[a]R^2=0.17; adjusted R^2=0.15; $F_{3, 238}$=15.95.

[b]Not applicable.

An ordinal logistic regression analysis was conducted to examine the relationship between the 3 predictor variables (ie, subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time) and the outcome variable (external IA symptoms; Table 10). The variables included in the ordinal logistic regression analysis met the assumptions of the statistical model—the dependent variable (ie, external IA symptoms) was treated as an ordinal variable, and the independent variables (ie, subjective happiness with nonessential internet use time, essential internet use time, and nonessential internet use time) were treated as continuous variables. The assumption of no multicollinearity was assessed by examining the VIFs, and the result (VIFs<1.1) indicated no multicollinearity. The results revealed that all 3 predictors were associated with changes in the external IA symptoms. Nonessential internet use time was associated with an increase in external IA symptoms (OR 1.19, 95% CI 1.03-1.39; *P*=.02). Essential internet use time was associated with an increase in external IA symptoms (OR 1.24, 95% CI 1.05-1.46; *P*=.01). Similarly, a higher unhappiness with nonessential internet use time was associated with increased external IA symptoms (OR 1.75, 95% CI 1.28-2.44; *P*<.001).

Table 10. Ordinal logistic regression predicting external internet addiction symptoms in middle adolescents.

Predictor	OR[a] (95% CI; SE)	*P* value
Subjective happiness with nonessential internet use time	1.75 (1.28-2.44; 0.56)	<.001
Essential internet use time	1.24 (1.05-1.46; 0.21)	.02
Nonessential internet use time	1.19 (1.03-1.39; 0.18)	.01

[a]OR: odds ratio.

A binary logistic regression was performed to assess the impact of subjective happiness with nonessential internet use time on the likelihood of being a dependent internet user (addicted) versus a nondependent internet user (unaddicted). The reference category was coded as class 1, representing nondependent internet users. The OR of 0.66 (95% CI 0.47-0.92) indicated the effect size. The *F*-measure was 75%. The results indicated a significant association between subjective happiness with nonessential internet use time and IA (*P*=.01), suggesting that higher levels of unhappiness were associated with a decreased likelihood of being unaddicted to the internet.

Discussion

Principal Findings

This study aimed to enhance our understanding of adolescent IA by addressing 3 key RQs. First, we examined the significance of differentiating between essential and nonessential internet use in relation to IA, offering a more nuanced view of IA beyond that of total screen time that is found in previous research. Second, we investigated whether subjective happiness linked to time spent on nonessential internet use could predict IA in adolescents. Finally, by studying early and middle adolescents separately, we identified age-specific differences in how time spent on essential and nonessential internet use and the related happiness contribute to IA.

Consistent with the extensive existing literature on IA, our study demonstrated that excessive engagement in nonessential internet activities, such as social media and online gaming, positively correlate with IA overall [62-64]. Furthermore, in our study, essential internet use did not exhibit a significant correlation with IA (*P*=.94 for early adolescents and *P*=.06 for middle adolescents), suggesting that essential internet activities such as remote learning or using the internet for school-required work may not carry the same level of addiction risk. This is similar to the results obtained in other studies [65,66]; Salubi et al [65] found that there was no correlation between essential internet use and IA among 390 surveyed university students in South Africa, whereas Pjevac et al [66] found that IA was lowest among adolescents who mostly used the internet for school purposes. This distinction of the purposes of internet use enables the development of targeted interventions and strategies to effectively address this issue. As digital technology becomes increasingly integral to tasks that once operated without it, acknowledging this technological shift is essential in our evolving digital society. It is necessary to recognize the evolving landscape of digital technology use and interpret digital technology use with caution, as emphasized by Squire and Steinkuehler [67], before delving into the complexities of IA. This caution entails considering the specific activities carried out on the internet, the motivations driving them, and the duration for which they occur.

Our study highlights similarities and key differences in IA between early and middle adolescents as our results revealed that, among early adolescents, nonessential internet use significantly predicted IA and its internal symptoms (related to the self). However, middle adolescents exhibited a more complex relationship with IA as both nonessential and essential internet use, in addition to unhappiness with nonessential internet activities, significantly predicted IA and its external symptoms (related to others).

Our results suggest that early adolescents are more likely to develop IA and experience internal IA symptoms as a result of nonessential internet use, whereas this influence seems to decrease as they get older. This finding is in line with the results of a meta-analysis (of 20 studies) on IA in adolescents reporting that IA is inversely proportional to age [57]. This difference suggests that the impact of nonessential digital technology use on IA may vary according to the developmental stage of adolescents. One possible explanation is that early adolescents are actively navigating their new identity and seeking peer acceptance, which can make them more susceptible to excessive use of digital technology [68]. The pressure to conform to peers' behaviors and expectations may override their ability to exercise self-control and resist the allure of digital technology [69]. Furthermore, in early adolescents, only nonessential internet use increased the likelihood of external IA symptoms. Our results are similar to those of a recent study that examined sociodemographic factors related to IA among a sample of 1664 adolescents in Serbia. The study reported the lowest IA rates among the younger adolescents and among those who used the internet for the purpose of schoolwork [66]. Our results could be interpreted as early adolescents not facing the same academic pressures and responsibilities that drive older adolescents to the excessive use of the internet for essential activities. Instead, early adolescent IA is more related to recreational online activities, underscoring the importance of tailoring preventive interventions to focus on nonessential use, such as social media and gaming, to equip early adolescents to develop healthier digital habits. Parental involvement and supervision play a vital role in preventing IA in early adolescents. Actively engaging with children, establishing open communication channels, and monitoring their digital technology use can create a supportive environment and set boundaries for healthy digital behaviors [27,70]. Emphasizing balance is also essential, highlighting the need for a well-rounded lifestyle that includes academic tasks, physical activities, face-to-face interactions, and nondigital recreational pursuits [71]. Self-control in early adolescents may be enhanced through self-control–training activities and participation in self-control–promoting programs [72].

In contrast to the results for early adolescents, our results demonstrated a more complex relationship between internet use and IA in middle adolescents. Our study found that, while nonessential internet use remains a significant predictor of IA, subjective happiness with nonessential internet activities also plays a role. Specifically, the unhappier middle adolescents were with the time spent on nonessential internet use, the more likely they were to experience IA. These findings are consistent with the findings of a similar study that surveyed a sample of adolescents and young adults (aged 15-24 years) and adults (aged 25-64 years), where dissatisfaction with smartphone use was linked to higher problematic internet use [73]. Our results contribute to the literature as there is a lack of research examining subjective happiness with digital technology use in relation to IA in adolescents.

This can be attributed to the increased academic pressure and the need to manage uncertainty, which drive middle adolescents to spend more time on essential internet activities, such as online educational tasks. The combination of information overload, academic demands, and the desire to avoid having to deal with uncertainty may lead to excessive internet use, further exacerbating IA.

Moreover, middle adolescents exhibited a broader range of factors influencing their external IA symptoms.

While, in early adolescents, only nonessential internet use increased the likelihood of external IA symptoms, in middle adolescents, both nonessential and essential internet use, as well as unhappiness with the time spent on nonessential internet use, increased external IA symptoms. This can be attributed to the intensified academic pressure in combination with information overload, leading them to neglect other aspects of their lives and driving them to the excessive use of the internet to meet academic demands [74,75]. The combination of these factors heightens the risk of IA [76-78]. Furthermore, these academic pressures can further impact social interactions as middle adolescents prioritize academic performance, leading them to show signs of stress in their engagement with others [74,79], manifesting in external symptoms of IA, as observed in our study. Recognizing the unique challenges faced by adolescents at different developmental stages is crucial for effectively tailoring preventive measures and interventions.

While our study focused on examining individual factors related to IA, considering factors related to the specific context in which our results were obtained should not be overlooked and could provide a more holistic understanding of IA among adolescents. Adolescents' internet use and their susceptibility to IA are influenced by cultural values, societal norms, and regional habits. For instance, in the Middle East, where our study was conducted, cultural expectations regarding academic achievement and family roles may contribute to heightened stress and pressure on adolescents [80,81], leading them to spend more time on online activities such as social media and gaming. The emphasis on academic success in our context combined with limited outdoor recreational opportunities due to the extremely hot climate might be driving adolescents to spend more time indoors and engage in online recreational activities, increasing their risk of IA.

In addition, the rapid digital transformation in the Middle East has created new challenges and opportunities for adolescents [82,83]. The increased availability of smartphones and internet access as well as a growing digital culture have made it easier for people to spend long hours on the web. While there are positive aspects of digitalization, it also raises concerns about the potential for excessive use and the development of IA. The addition of environmental factors such as the local affluent culture of convenience, the wide availability of digital devices even at very young ages, and the wide presence of high-speed

internet further influences how adolescents interact with digital technology.

Our study has important implications for the field of IA and for promoting healthier digital technology use habits in adolescents. The findings suggest that interventions aimed at addressing IA should focus on nonessential use specifically, particularly in early adolescents, rather than considering internet use as a whole. Because IA is a complex outcome influenced by multiple physical, psychological, and technological factors rather than a single issue, this targeted approach can help effectively address the factors contributing to IA while considering the developmental factors of adolescents. While there is a dearth of data from the region, there is evidence from a systematic literature review on policy and prevention approaches for disordered gaming and internet use suggesting that, due to the global prevalence of the phenomena and the geographically dispersed nature of preventative programs, it is imperative that prevention efforts be integrated across national borders while taking cultural differences into consideration [84].

The following recommendations are a recapitulation of the evidence produced by the limited international reviews that have examined the effectiveness of programs and interventions in the prevention of IA [71,84-87].

Early identification and intervention are key to addressing IA among adolescents. Schools can implement evidence-based prevention programs specifically targeting IA. These programs can incorporate educational components, skill-building exercises, and peer support networks to help adolescents develop healthy online habits and balance their time between essential and nonessential activities [87]. Comprehensive school-based programs are essential to effectively address IA. These programs should focus on increasing awareness, develop coping strategies, and promoting responsible digital technology use. These programs can be integrated into the school curricula and should include methods to improve self-control and enhance knowledge of the underlying processes that contribute to excessive online activity, such as constant notifications [88,89]. In addition, schools can implement practical digital well-being initiatives, such as mindfulness exercises and the integration of technology breaks into the daily schedule, to promote healthier digital habits. Moreover, promoting digital literacy and responsible digital technology use is crucial for educating adolescents on the potential risks and consequences associated with excessive digital technology use [90]. Given that adolescents' behaviors are largely shaped by other external factors such as their families and external environments [21], schools play a crucial role in extending these preventive programs to parents and caregivers. Understanding children's intentions when interacting with the internet is crucial and emphasizes the need for smarter digital parenting tools that provide enhanced monitoring and guidance. A healthier digital environment can be created by considering intentionality and using smart tools. Devices and programs should be designed to isolate essential and nonessential uses, advising children to avoid installing or visiting purely nonessential social media applications and sites on the same device in which they engage in essential activities.

These recommendations must be supported by designing and enacting policies. Policy makers can design programs targeting the broader environment in which adolescents live, addressing the underlying factors contributing to maladaptive coping, such as reducing academic pressure and improving family support. By reducing dependence on nonessential internet use, these policies, in turn, can empower adolescents to navigate their digital environments more responsibly.

Certain limitations of this study must be considered when interpreting the results. First, due to the cross-sectional design of our study, the generalizability of our findings is limited. Another limitation of our study was that the responses provided by the adolescents were based on self-report, potentially introducing reporting bias. It is possible that the adolescents did not accurately report the time spent on essential and nonessential activities and symptoms of IA. To mitigate the influence of reporting bias, future studies should incorporate objective measures, such as using data on actual screen time, instead of relying solely on self-reported information. Another limitation of this study is that the context of essential and nonessential internet use could have been interpreted by adolescents based on their subjective experiences and perceptions. For example, spending time with friends on social media during the examination period may be considered essential for adolescents who want to receive moral support from their colleagues. Nevertheless, our study left it at the discretion of the participants. The impact of the specific activity or content that adolescents consume or spend excessive time on and the effect it has on them should be scrutinized in more detail and is an area that requires attention. Our sample's gender imbalance, which was characterized by a higher proportion of female participants, may restrict the generalizability of the results to both genders. Furthermore, we adhered to sample size estimates using the method by Green [49] to guarantee sufficient statistical power for our study. The deliberate choice not to segregate the sample by gender was made to ensure that the sample size remained sufficient for obtaining consistent and dependable regression results. The male sample was relatively too small to allow for that option. Although the existing literature on gender differences in IA is inconclusive, with research presenting inconsistent findings [91-95], it is important to exercise caution when interpreting the results of this study due to the higher proportion of female participants. Future studies should aim for a more even gender distribution to comprehensively investigate any disparities in IA between male and female individuals.

The strength of our study lies in its inclusion of a non-WEIRD population, specifically focusing on adolescent IA in the Middle East, where limited research on this area has been conducted. Another strength of our study is that it addresses the research gap by examining both essential and nonessential internet use in the same sample. While previous research has explored the different online activities that contribute to IA [65,96], few studies have specifically investigated the impact of essential and nonessential internet use within a single study. Another strength of our study lies in the separate analysis of IA factors in early and middle adolescents, which allowed us to capture the unique dynamics and influences shaping IA in each age

group, thus enhancing the understanding of IA in relation to developmental stages.

Conclusions

In conclusion, this study sheds light on the complex relationship between digital technology use and IA among early and middle adolescents. For many years, screen time has been widely used as a prominent measure to assess the extent of IA among adolescents. As the digital society continues to evolve rapidly, relying on screen time as a stand-alone criterion fails to provide a comprehensive understanding of phenomena such as IA. Therefore, it is imperative to distinguish between essential and nonessential digital technology use, acknowledging diverse activities and their implications for a better understanding of individuals' engagement with digital technology. The findings of our study provide valuable insights for researchers, practitioners, and policy makers in addressing IA among adolescents. By recognizing the distinct nature of internet activities and considering developmental factors, effective interventions and strategies can be developed to promote healthy digital technology use and prevent IA in adolescents.

Acknowledgments

"Technology Overuse Amongst Adolescents in Qatar: current status, challenges, and recommendations for policy" is a joint project among 3 members of the Qatar Foundation for Education, Science, and Community Development; the Doha International Family Institute; the World Innovation Summit for Education; and the Word Innovation Summit for Health. The statements made herein are solely the responsibility of the authors and do not reflect the statements or opinions of these institutions.

Authors' Contributions

KC conceptualized the study, prepared and analyzed the data, and wrote the paper. MA helped with data preparation and verified the statistical analysis. AA, SA-H, and AB participated in the study conceptualization and design and conducted the data collection. DAH critically reviewed and revised the Conclusions section in the paper. RA participated in all stages and supervised the research. All authors reviewed the paper and provided feedback. All authors approved the final version of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey design.
[DOCX File , 608 KB - humanfactors_v12i1e62955_app1.docx]

References

1. Vogels EA, Gelles-Watnick R, Massarat N. Teens, social media and technology 2022. Pew Research Center. 2022 Aug 10. URL: https://www.pewresearch.org/internet/2022/08/10/teens-social-media-and-technology-2022/ [accessed 2023-05-31]
2. Balhara YP, Mahapatra A, Sharma P, Bhargava R. Problematic internet use among students in South-East Asia: current state of evidence. Indian J Public Health 2018;62(3):197-210 [FREE Full text] [doi: 10.4103/ijph.IJPH_288_17] [Medline: 30232969]
3. Chung TW, Sum SM, Chan MW. Adolescent internet addiction in Hong Kong: prevalence, psychosocial correlates, and prevention. J Adolesc Health 2019 Jun;64(6S):S34-S43. [doi: 10.1016/j.jadohealth.2018.12.016] [Medline: 31122547]
4. Lee JY, Kim SY, Bae KY, Kim JM, Shin IS, Yoon JS, et al. Prevalence and risk factors for problematic internet use among rural adolescents in Korea. Asia Pac Psychiatry 2018 Jun;10(2):e12310. [doi: 10.1111/appy.12310] [Medline: 29356370]
5. Xin M, Xing J, Pengfei W, Houru L, Mengcheng W, Hong Z. Online activities, prevalence of internet addiction and risk factors related to family and school among adolescents in China. Addict Behav Rep 2018 Jun;7:14-18 [FREE Full text] [doi: 10.1016/j.abrep.2017.10.003] [Medline: 29450251]
6. Oh WO. Factors influencing internet addiction tendency among middle school students in Gyeong-buk area. Taehan Kanho Hakhoe Chi 2003 Dec;33(8):1135-1144. [doi: 10.4040/jkan.2003.33.8.1135] [Medline: 15314360]
7. Morrison CM, Gore H. The relationship between excessive internet use and depression: a questionnaire-based study of 1,319 young people and adults. Psychopathology 2010;43(2):121-126. [doi: 10.1159/000277001] [Medline: 20110764]
8. Singh N, Barmola KC. Internet addiction, mental health and academic performance of school students/adolescents. Int J Indian Psychol 2015 Jun 25;2(3). [doi: 10.25215/0203.053]
9. Tateno M, Teo AR, Ukai W, Kanazawa J, Katsuki R, Kubo H, et al. Internet addiction, smartphone addiction, and hikikomori trait in Japanese young adult: social isolation and social network. Front Psychiatry 2019;10:455 [FREE Full text] [doi: 10.3389/fpsyt.2019.00455] [Medline: 31354537]
10. Young KS. Internet addiction: the emergence of a new clinical disorder. Cyberpsychol Behav 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]

11. Shaw M, Black DW. Internet addiction: definition, assessment, epidemiology and clinical management. CNS Drugs 2008;22(5):353-365. [doi: 10.2165/00023210-200822050-00001] [Medline: 18399706]
12. Young KS, de Abreu CN. Internet Addiction: A Handbook and Guide to Evaluation and Treatment. Hoboken, NJ: John Wiley & Sons; 2007.
13. Mihajlov M, Vejmelka L. Internet addiction: a review of the first twenty years. Psychiatr Danub 2017 Sep 20;29(3):260-272 [FREE Full text] [doi: 10.24869/psyd.2017.260] [Medline: 28949307]
14. Rad MS, Martingano AJ, Ginges J. Toward a psychology of Homo sapiens: making psychological science more representative of the human population. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11401-11405. [doi: 10.1073/pnas.1721165115] [Medline: 30397114]
15. Schaller M, Norenzayan A, Heine SJ, Yamagishi T, Kameda T. Evolutionary Psychology Meets Cultural Psychology: Review of Evolution, Culture and the Human Mind. East Sussex, UK: Psychology Press; 2009.
16. Wigfield A, Byrnes JP, Eccles JS. Development during early and middle adolescence. In: Alexander PA, Winne PH, editors. Handbook of Educational Psychology. New York, NY: Routledge; 2006.
17. Kemp S. Digital 2023: Qatar. Data Reportal. 2023 Feb 13. URL: https://datareportal.com/reports/digital-2023-qatar [accessed 2023-11-08]
18. World Bank open data. World Bank Group. URL: https://data.worldbank.org/ [accessed 2024-08-10]
19. Saquib J. Internet addiction among Saudi Arabian youth. Int J Health Sci (Qassim) 2020;14(2):1-2 [FREE Full text] [Medline: 32206053]
20. Al-Khani AM, Saquib J, Rajab AM, Khalifa MA, Almazrou A, Saquib N. Internet addiction in Gulf countries: a systematic review and meta-analysis. J Behav Addict 2021 Sep 06;10(3):601-610 [FREE Full text] [doi: 10.1556/2006.2021.00057] [Medline: 34491902]
21. Chemnad K, Aziz M, Abdelmoneium AO, Al-Harahsheh S, Baghdady A, Al Motawaa FY, et al. Adolescents' internet addiction: does it all begin with their environment? Child Adolesc Psychiatry Ment Health 2023 Jul 04;17(1):87 [FREE Full text] [doi: 10.1186/s13034-023-00626-7] [Medline: 37403108]
22. Ko D, Yao M. 11. Internet addiction: an cross-cultural perspective. In: The Psychology of Social Networking Vol.2: Identity and Relationships in Online Communities. Warsaw, Poland: De Gruyter Open Poland; 2016:141-158.
23. Aşut Ö, Abuduxike G, Acar-Vaizoğlu S, Cali S. Relationships between screen time, internet addiction and other lifestyle behaviors with obesity among secondary school students in the Turkish Republic of Northern Cyprus. Turk J Pediatr 2019 Aug 25;61(4):568-579 [FREE Full text] [doi: 10.24953/turkjped.2019.04.014] [Medline: 31990475]
24. Twenge JM, Campbell WK. Associations between screen time and lower psychological well-being among children and adolescents: evidence from a population-based study. Prev Med Rep 2018 Dec;12:271-283 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.003] [Medline: 30406005]
25. Lissak G. Adverse physiological and psychological effects of screen time on children and adolescents: literature review and case study. Environ Res 2018 Jul;164:149-157. [doi: 10.1016/j.envres.2018.01.015] [Medline: 29499467]
26. Sousa Tavares HL, Souza Costa D, Soares AA, Kestelman I, da Silva AG, Malloy-Diniz LF, et al. Gaming addiction and screen time in a context of increase of internalizing symptoms: moderation evaluation. Clin Child Psychol Psychiatry 2023 Jan 03;28(1):212-223. [doi: 10.1177/13591045221125329] [Medline: 36062317]
27. Chemnad K, Alshakhsi S, Al-Harahsheh S, Abdelmoneium AO, Al-Khalaf MS, Baghdady A, et al. Is it contagious? Does parents' internet addiction impact their adolescents' internet addiction? Soc Sci Comput Rev 2022 Aug 05;41(5):1691-1711. [doi: 10.1177/08944393221117408]
28. Cerniglia L, Zoratto F, Cimino S, Laviola G, Ammaniti M, Adriani W. Internet addiction in adolescence: neurobiological, psychosocial and clinical issues. Neurosci Biobehav Rev 2017 May;76(Pt A):174-184. [doi: 10.1016/j.neubiorev.2016.12.024] [Medline: 28027952]
29. Yen JY, Ko CH, Yen CF, Wu HY, Yang MJ. The comorbid psychiatric symptoms of internet addiction: attention deficit and hyperactivity disorder (ADHD), depression, social phobia, and hostility. J Adolesc Health 2007 Jul;41(1):93-98. [doi: 10.1016/j.jadohealth.2007.02.002] [Medline: 17577539]
30. Costescu C, Chelba I, Ro an A, Kovari A, Katona J. Cognitive patterns and coping mechanisms in the context of internet use. Appl Sci 2021 Feb 01;11(3):1302. [doi: 10.3390/app11031302]
31. Chou WJ, Huang MF, Chang YP, Chen YM, Hu HF, Yen CF. Social skills deficits and their association with internet addiction and activities in adolescents with attention-deficit/hyperactivity disorder. J Behav Addict 2017 Mar 01;6(1):42-50 [FREE Full text] [doi: 10.1556/2006.6.2017.005] [Medline: 28245666]
32. Ciccia AH, Meulenbroek P, Turkstra LS. Adolescent brain and cognitive developments: implications for clinical assessment in traumatic brain injury. Top Lang Disord 2009;29(3):249-265 [FREE Full text] [doi: 10.1097/TLD.0b013e3181b53211] [Medline: 30220763]
33. Shah JY, Gardner WL. Handbook of Motivation Science. New York, NY: Guilford Publications; 2008.
34. Koç P. Internet addiction and subjective wellbeing in university students. Journal of Positive Psychology and Wellbeing 2017;1(1):34-41 [FREE Full text]
35. Fauzi MA, Harun SA, Martin T, Paiman N, Hussin MS. Smartphone addiction among young adolescents: integration of flow theory and fear of missing out. Int J Inf Manag Sci 2021;32:39-58 [FREE Full text]

36. Harter S. Emerging self-processes during childhood and adolescence. In: Leary MR, Tangney JP, editors. Handbook of Self and Identity. New York, NY: Guilford Press; 2012:680-715.
37. Plancherel B, Bolognini M, Halfon O. Coping strategies in early and mid-adolescence. Eur Psychol 1998 Sep;3(3):192-201. [doi: 10.1027/1016-9040.3.3.192]
38. Blyth DA, Traeger CM. The self‐concept and self‐esteem of early adolescents. Theory Pract 2009 Nov 05;22(2):91-97. [doi: 10.1080/00405848309543045]
39. Brinthaupt TM, Lipka RP. Understanding Early Adolescent Self and Identity: Applications and Interventions. New York, NY: State University of New York Press; 2002.
40. Scheiner C, Grashoff J, Kleindienst N, Buerger A. Mental disorders at the beginning of adolescence: prevalence estimates in a sample aged 11-14 years. Public Health Pract (Oxf) 2022 Dec;4:100348 [FREE Full text] [doi: 10.1016/j.puhip.2022.100348] [Medline: 36545674]
41. Joshi P, Shukla S. Middle childhood and adolescence: development and learning. In: Child Development and Education in the Twenty-First Century. Singapore, Singapore: Springer; 2019.
42. Umberson D, Crosnoe R, Reczek C. Social relationships and health behavior across life course. Annu Rev Sociol 2010 Aug 01;36(1):139-157 [FREE Full text] [doi: 10.1146/annurev-soc-070308-120011] [Medline: 21921974]
43. Borghuis J, Denissen JJ, Oberski D, Sijtsma K, Meeus WH, Branje S, et al. Big Five personality stability, change, and codevelopment across adolescence and early adulthood. J Pers Soc Psychol 2017 Oct;113(4):641-657 [FREE Full text] [doi: 10.1037/pspp0000138] [Medline: 28253001]
44. Havewala M, Bowker JC, Smith KA, Rose-Krasnor L, Booth-LaForce C, Laursen B, et al. Peer influence during adolescence: the moderating role of parental support. Children (Basel) 2021 Apr 17;8(4):306 [FREE Full text] [doi: 10.3390/children8040306] [Medline: 33920622]
45. Ren Y, Chi X, Bu H, Huang L, Wang S, Zhang Y, et al. Warm and harsh parenting, self-kindness and self-judgment, and well-being: an examination of developmental differences in a large sample of adolescents. Children (Basel) 2023 Feb 19;10(2):406 [FREE Full text] [doi: 10.3390/children10020406] [Medline: 36832535]
46. Dienlin T, Johannes N. The impact of digital technology use on adolescent well-being . Dialogues Clin Neurosci 2022 Apr 01;22(2):135-142. [doi: 10.31887/dcns.2020.22.2/tdienlin]
47. Moreno MA, Binger K, Zhao Q, Eickhoff J. Measuring interests not minutes: development and validation of the Adolescents' Digital Technology Interactions and Importance scale (ADTI). J Med Internet Res 2020 Feb 12;22(2):e16736 [FREE Full text] [doi: 10.2196/16736] [Medline: 32049068]
48. Aziz M, Chemnad K, Al-Harahsheh S, Abdelmoneium AO, Bagdady A, Hassan DA, et al. The influence of adolescents essential and non-essential use of technology and internet addiction on their physical and mental fatigues. Sci Rep 2024 Jan 19;14(1):1745 [FREE Full text] [doi: 10.1038/s41598-024-51655-x] [Medline: 38242916]
49. Green SB. How many subjects does it take to do a regression analysis. Multivariate Behav Res 1991 Jul 01;26(3):499-510. [doi: 10.1207/s15327906mbr2603_7] [Medline: 26776715]
50. Schönbrodt FD, Perugini M. At what sample size do correlations stabilize? J Res Pers 2013 Oct;47(5):609-612. [doi: 10.1016/j.jrp.2013.05.009]
51. Smetana JG, Campione-Barr N, Metzger A. Adolescent development in interpersonal and societal contexts. Annu Rev Psychol 2006 Jan;57(1):255-284. [doi: 10.1146/annurev.psych.57.102904.190124] [Medline: 16318596]
52. Brislin RW. Back-translation for cross-cultural research. J Cross Cult Psychol 1970 Sep 01;1(3):185-216. [doi: 10.1177/135910457000100301]
53. Lyons-Thomas J. Using think aloud protocols in the validity investigation of an assessment of complex thinking. University of British Columbia. 2014. URL: https://open.library.ubc.ca/soa/cIRcle/collections/ubctheses/24/items/1.0167403 [accessed 2025-01-16]
54. Strittmatter E, Brunner R, Fischer G, Parzer P, Resch F, Kaess M. [Association of peer victimization, coping, and pathological internet use among adolescents]. Z Kinder Jugendpsychiatr Psychother 2014 Mar;42(2):85-94. [doi: 10.1024/1422-4917/a000275] [Medline: 24571814]
55. Young K. Internet Addiction Test (IAT). Wood Dale, IL: Stoelting Company; 2016.
56. Ursachi G, Horodnic IA, Zait A. How reliable are measurement scales? External factors with indirect influence on reliability estimators. Procedia Econ Financ 2015;20:679-686. [doi: 10.1016/s2212-5671(15)00123-9]
57. Lozano-Blasco R, Latorre-Martínez M, Cortés-Pascual A. Screen addicts: a meta-analysis of internet addiction in adolescence. Child Youth Serv Rev 2022 Apr;135:106373. [doi: 10.1016/j.childyouth.2022.106373]
58. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: conducting a content validity study in social work research. Soc Work Res 2003 Jun 01;27(2):94-104. [doi: 10.1093/swr/27.2.94]
59. JASP - a fresh way to do statistics. JASP. URL: https://jasp-stats.org/ [accessed 2021-11-29]
60. Cureton EE, Mulaik SA. The weighted varimax rotation and the promax rotation. Psychometrika 1975 Jun;40(2):183-195. [doi: 10.1007/bf02291565]
61. Kaiser HF, Rice J. Little Jiffy, Mark IV. Educ Psychol Meas 1974 Apr 01;34(1):111-117. [doi: 10.1177/001316447403400115]

62. Ciacchini R, Orrù G, Cucurnia E, Sabbatini S, Scafuto F, Lazzarelli A, et al. Social media in adolescents: a retrospective correlational study on addiction. Children (Basel) 2023 Jan 31;10(2):278 [FREE Full text] [doi: 10.3390/children10020278] [Medline: 36832407]
63. Putra PY, Fithriyah I, Zahra Z. Internet addiction and online gaming disorder in children and adolescents during COVID-19 pandemic: a systematic review. Psychiatry Investig 2023 Mar;20(3):196-204 [FREE Full text] [doi: 10.30773/pi.2021.0311] [Medline: 36990662]
64. Kuss DJ, Kristensen AM, Lopez-Fernandez O. Internet addictions outside of Europe: a systematic literature review. Comput Hum Behav 2021 Feb;115:106621. [doi: 10.1016/j.chb.2020.106621]
65. Salubi O, Muchaonyerwa N, Nekhwevha F. Digital media usage and prevalence of internet addiction among undergraduate students in South Africa. Int J Web Appl 2018 Dec;10(4):127 [FREE Full text] [doi: 10.6025/ijwa/2018/10/4/127-136]
66. Ana P, Teodora S, Mirjana J. Sociodemographic factors related to internet addiction among adolescents in Serbia. Vojnosanitetski Pregled 2024;81(7):438-446. [doi: 10.2298/vsp240307037p]
67. Squire KD, Steinkuehler C. The problem with screen time. Teach Coll Rec 2017 Dec 01;119(12):1-24. [doi: 10.1177/016146811711901207]
68. Caravita SC, Sijtsema JJ, Rambaran JA, Gini G. Peer influences on moral disengagement in late childhood and early adolescence. J Youth Adolesc 2014 Feb 10;43(2):193-207. [doi: 10.1007/s10964-013-9953-1] [Medline: 23660831]
69. Steinberg L, Albert D, Cauffman E, Banich M, Graham S, Woolard J. Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. Dev Psychol 2008 Nov;44(6):1764-1778. [doi: 10.1037/a0012955] [Medline: 18999337]
70. Schmid E, Garrels V. Parental involvement and educational success among vulnerable students in vocational education and training. Educ Res 2021 Nov 15;63(4):456-473. [doi: 10.1080/00131881.2021.1988672]
71. Vondráčková P, Gabrhelík R. Prevention of internet addiction: a systematic review. J Behav Addict 2016 Dec;5(4):568-579 [FREE Full text] [doi: 10.1556/2006.5.2016.085] [Medline: 27998173]
72. Allemand M, Job V, Mroczek DK. Self-control development in adolescence predicts love and work in adulthood. J Pers Soc Psychol 2019 Sep;117(3):621-634 [FREE Full text] [doi: 10.1037/pspp0000229] [Medline: 30550330]
73. Alshakhsi S, Chemnad K, Almourad MB, Altuwairiqi M, McAlaney J, Ali R. Problematic internet usage: the impact of objectively recorded and categorized usage time, emotional intelligence components and subjective happiness about usage. Heliyon 2022 Oct;8(10):e11055 [FREE Full text] [doi: 10.1016/j.heliyon.2022.e11055] [Medline: 36281419]
74. Anniko MK, Boersma K, Tillfors M. Sources of stress and worry in the development of stress-related mental health problems: a longitudinal investigation from early- to mid-adolescence. Anxiety Stress Coping 2019 Mar 22;32(2):155-167. [doi: 10.1080/10615806.2018.1549657] [Medline: 30465445]
75. Dissing AS, Jørgensen TB, Gerds TA, Rod NH, Lund R. High perceived stress and social interaction behaviour among young adults. A study based on objective measures of face-to-face and smartphone interactions. PLoS One 2019 Jul 26;14(7):e0218429 [FREE Full text] [doi: 10.1371/journal.pone.0218429] [Medline: 31348787]
76. Feng Y, Ma Y, Zhong Q. The relationship between adolescents' stress and internet addiction: a mediated-moderation model. Front Psychol 2019;10:2248 [FREE Full text] [doi: 10.3389/fpsyg.2019.02248] [Medline: 31636590]
77. Hofstede G. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. Thousand Oaks, CA: SAGE Publications; 2001.
78. Wu CY, Lee MB, Liao SC, Chang LR. Risk factors of internet addiction among internet users: an online questionnaire survey. PLoS One 2015 Oct 10;10(10):e0137506 [FREE Full text] [doi: 10.1371/journal.pone.0137506] [Medline: 26462196]
79. Pascoe MC, Hetrick SE, Parker AG. The impact of stress on students in secondary school and higher education. Int J Adolesc Youth 2019 Apr 11;25(1):104-112. [doi: 10.1080/02673843.2019.1596823]
80. Sellami A. Parental influence on student educational expectations: results from the 2012 Qatar education study. Int J High Educ 2019 Jul;8(4):189. [doi: 10.5430/ijhe.v8n4p189]
81. Ali J, Alsakhe H, Ibrahim I, Khattab N, Madeeha M, Shouia M. Accounting for educational expectations and achievement among native and migrant students in Qatar. Int Stud Sociol Educ 2022 Feb 15;33(2):159-178. [doi: 10.1080/09620214.2022.2034518]
82. Melanyina MV, Ruzina EI, Shkvarya LV, Verenikina AY. Digital transformation in the middle east: directions, outcomes and challenges. Economic Studies Journal 2024(4):35-47 [FREE Full text]
83. Marino C, Gini G, Vieno A, Spada MM. A comprehensive meta-analysis on problematic Facebook use. Comput Hum Behav 2018 Jun;83:262-277. [doi: 10.1016/j.chb.2018.02.009]
84. King DL, Delfabbro PH, Doh YY, Wu AM, Kuss DJ, Pallesen S, et al. Policy and prevention approaches for disordered and hazardous gaming and internet use: an international perspective. Prev Sci 2018 Feb 4;19(2):233-249. [doi: 10.1007/s11121-017-0813-1] [Medline: 28677089]
85. Király O, Griffiths MD, King DL, Lee HK, Lee SY, Bányai F, et al. Policy responses to problematic video game use: a systematic review of current measures and future possibilities. J Behav Addict 2018 Sep 01;7(3):503-517 [FREE Full text] [doi: 10.1556/2006.6.2017.050] [Medline: 28859487]
86. Bağatarhan T, Siyez DM. Programs for preventing internet addiction during adolescence: a systematic review. Addicta 2015 May 31;4:243-265. [doi: 10.15805/addicta.2017.4.2.0015]

87. Throuvala MA, Griffiths MD, Rennoldson M, Kuss DJ. School-based prevention for adolescent internet addiction: prevention is the key. A systematic literature review. Curr Neuropharmacol 2019 May 09;17(6):507-525. [doi: 10.2174/1570159x16666180813153806]
88. Gazzaley A, Rosen LD. The Distracted Mind: Ancient Brains in a High-Tech World. Cambridge, MA: MIT Press; 2016.
89. Liao M, Sundar SS. Sound of silence: does muting notifications reduce phone use? Comput Hum Behav 2022 Sep;134:107338. [doi: 10.1016/j.chb.2022.107338]
90. Hyman A, Stewart K, Jamin AM, Novak Lauscher H, Stacy E, Kasten G, et al. Testing a school-based program to promote digital health literacy and healthy lifestyle behaviours in intermediate elementary students: the Learning for Life program. Prev Med Rep 2020 Sep;19:101149 [FREE Full text] [doi: 10.1016/j.pmedr.2020.101149] [Medline: 32670779]
91. Mo PK, Chan VW, Wang X, Lau JT. Gender difference in the association between internet addiction, self-esteem and academic aspirations among adolescents: a structural equation modelling. Comput Educ 2020 Oct;155:103921. [doi: 10.1016/j.compedu.2020.103921]
92. Chi X, Hong X, Chen X. Profiles and sociodemographic correlates of internet addiction in early adolescents in southern China. Addict Behav 2020 Jul;106:106385. [doi: 10.1016/j.addbeh.2020.106385] [Medline: 32187574]
93. Peris M, de la Barrera U, Schoeps K, Montoya-Castilla I. Psychological risk factors that predict social networking and internet addiction in adolescents. Int J Environ Res Public Health 2020 Jun 26;17(12):4598 [FREE Full text] [doi: 10.3390/ijerph17124598] [Medline: 32604834]
94. Li Q, Dai W, Zhong Y, Wang L, Dai B, Liu X. The mediating role of coping styles on impulsivity, behavioral inhibition/approach system, and internet addiction in adolescents from a gender perspective. Front Psychol 2019 Oct 24;10:2402 [FREE Full text] [doi: 10.3389/fpsyg.2019.02402] [Medline: 31708840]
95. Peterka-Bonetta J, Sindermann C, Elhai JD, Montag C. Personality associations with smartphone and internet use disorder: a comparison study including links to impulsivity and social anxiety. Front Public Health 2019 Jun 11;7:127 [FREE Full text] [doi: 10.3389/fpubh.2019.00127] [Medline: 31245341]
96. Seo M, Kang HS, Yom YH. Internet addiction and interpersonal problems in Korean adolescents. Comput Inform Nurs 2009;27(4):226-233. [doi: 10.1097/ncn.0b013e3181a91b3f]

Abbreviations

IA: internet addiction
IADQ: Internet Addiction Diagnostic Questionnaire
KMO: Kaiser-Meyer-Olkin
OR: odds ratio
RQ: research question
VIF: variance inflation factor
WEIRD: Western, educated, industrialized, rich, and democratic

Edited by A Kushniruk; submitted 05.06.24; peer-reviewed by A Rizvi, S Ostovar, T Safiye; comments to author 31.07.24; revised version received 17.09.24; accepted 01.12.24; published 10.02.25.

Please cite as:
Chemnad K, Aziz M, Al- Harahsheh S, Abdelmoneium A, Baghdady A, Alsayed Hassan D, Ali R
Assessing the Relationship Between the Type of Internet Use and Internet Addiction in Early and Middle Adolescents: Cross-Sectional Study From Qatar
JMIR Hum Factors 2025;12:e62955
URL: https://humanfactors.jmir.org/2025/1/e62955
doi:10.2196/62955
PMID:

Original Paper

The Effects of Presenting AI Uncertainty Information on Pharmacists' Trust in Automated Pill Recognition Technology: Exploratory Mixed Subjects Study

Jin Yong Kim[1], BSE; Vincent D Marshall[2], MS; Brigid Rowell[2], MA; Qiyuan Chen[1], MSE; Yifan Zheng[2], PharmD; John D Lee[3], PhD; Raed Al Kontar[1], PhD; Corey Lester[2], PharmD, PhD; Xi Jessie Yang[1], PhD

[1]Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, United States

[2]College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

[3]Industrial and Systems Engineering, University of Wisconsin-Madison, Madison, MI, United States

Corresponding Author:
Xi Jessie Yang, PhD
Industrial and Operations Engineering
University of Michigan
1640 IOE
1205 Beal Avenue
Ann Arbor, MI, 48105
United States
Phone: 1 7347630541
Email: xijyang@umich.edu

Abstract

Background: Dispensing errors significantly contribute to adverse drug events, resulting in substantial health care costs and patient harm. Automated pill verification technologies have been developed to aid pharmacists with medication dispensing. However, pharmacists' trust in such automated technologies remains unexplored.

Objective: This study aims to investigate pharmacists' trust in automated pill verification technology designed to support medication dispensing.

Methods: Thirty licensed pharmacists in the United States performed a web-based simulated pill verification task to determine whether an image of a filled medication bottle matched a known reference image. Participants completed a block of 100 verification trials without any help, and another block of 100 trials with the help of an imperfect artificial intelligence (AI) aid recommending acceptance or rejection of a filled medication bottle. The experiment used a mixed subjects design. The between-subjects factor was the AI aid type, with or without an AI uncertainty plot. The within-subjects factor was the four potential verification outcomes: (1) the AI rejects the incorrect drug, (2) the AI rejects the correct drug, (3) the AI approves the incorrect drug, and (4) the AI approves the correct drug. Participants' trust in the AI system was measured. Mixed model (generalized linear models) tests were conducted with 2-tailed *t* tests to compare the means between the 2 AI aid types for each verification outcome.

Results: Participants had an average trust propensity score of 72 (SD 18.08) out of 100, indicating a positive attitude toward trusting automated technologies. The introduction of an uncertainty plot to the AI aid significantly enhanced pharmacists' end trust (t_{28}=–1.854; P=.04). Trust dynamics were influenced by AI aid type and verification outcome. Specifically, pharmacists using the AI aid with the uncertainty plot had a significantly larger trust increment when the AI approved the correct drug ($t_{78.98}$=3.93; P<.001) and a significantly larger trust decrement when the AI approved the incorrect drug ($t_{2939.72}$=–4.78; P<.001). Intriguingly, the absence of the uncertainty plot led to an increase in trust when the AI correctly rejected an incorrect drug, whereas the presence of the plot resulted in a decrease in trust under the same circumstances ($t_{509.77}$=–3.96; P<.001). A pronounced "negativity bias" was observed, where the degree of trust reduction when the AI made an error exceeded the trust gain when the AI made a correct decision (z=–11.30; P<.001).

Conclusions: To the best of our knowledge, this study is the first attempt to examine pharmacists' trust in automated pill verification technology. Our findings reveal that pharmacists have a favorable disposition toward trusting automation. Moreover, providing uncertainty information about the AI's recommendation significantly boosts pharmacists' trust in AI aid, highlighting the importance of developing transparent AI systems within health care.

(JMIR Hum Factors 2025;12:e60273) doi:10.2196/60273

KEYWORDS

artificial intelligence; human-computer interaction; uncertainty communication; visualization; medication errors; safety; artificial intelligence aid; pharmacists; pill verification; automation

Introduction

Background

Pharmacists play a pivotal role in ensuring patients receive the correct medications as prescribed by health care providers. This involves a critical verification task, where pharmacists must match the medication dispensed in filled bottles with the prescription labels. Dispensing errors, defined as instances when patients receive the wrong drug or dosage, significantly contribute to preventable adverse drug events, leading to approximately 700,000 emergency department visits and 100,000 hospital admissions each year [1]. Several challenges contribute to these errors including, but not limited to, limitations in current technology; lack of standardized procedures; and the high cognitive workload imposed on pharmacy staff, who often manage numerous tasks simultaneously [1-4]. To enhance patient health outcomes, reduce unnecessary health care costs, and alleviate the burden on pharmacists, it is essential to develop and implement reliable tools that minimize the risk of dispensing errors [5].

Since the 1990s, the implementation of barcode scanning systems has been advocated as a means to reduce medication errors [6]. The adoption of such systems within pharmacies and broader health care environments has led to a notable reduction in medication errors [7-10]. Nevertheless, research indicates that barcode scanning systems are not immune to workarounds and human errors [11-13]. Moreover, these systems do not adequately address the challenges faced by overburdened pharmacists [14-17].

In response to these challenges, pill counting and verification or recognition systems using image classification technologies have emerged [18-21]. Innovations like Eyecon and VIVID use vision-based methods to count medications placed on the tray. More recently, advancements have been made in automated pill identification through feature engineering. For example, Yu et al [22] and Yu et al [23] proposed an automatic pill recognition method based on pill imprints, achieving an accuracy of 86.01% and 90.46%, respectively. Caban et al [18] used a modified shape distribution technique to determine the shape, color, and imprint of a pill to identify the drug. The proposed technique was evaluated with 568 of the most prescribed drugs in the United States and achieved a 91.13% accuracy.

The advent of deep learning has further enhanced the capabilities of automated pill recognition systems [5,24]. For instance, Larios Delgado et al [5] developed a pill recognition method using 2 deep learning models. They used a deep convolutional neural network model for pill blob detection to isolate the pill from the background and then passed the output to a deep learning–based classifier to identify the 5 most likely pills with 94% accuracy [5]. Similarly, Wong et al [25] proposed a deep convolutional network model and achieved a mean accuracy of 95.35%. Lester et al [24] trained a ResNet-18 deep learning model to predict the labeled features of a medication product using an image showing medication inside a filled prescription bottle. In a test set containing 65,274 images of 345 unique oral drug products, the overall macroaverage precision, that is, the mean of precision values for each class, was 98.5%.

Despite the impressive strides in model accuracy, realizing the potential of these technologies is only possible if people establish appropriate trust in them. Trust in automation, defined as "the attitude that an agent will help achieve an individual's goals in situations characterized by uncertainty and vulnerability" [26], is one of the most crucial factors determining the use of automation [27,28]. There is a growing body of research examining people's trust in autonomous and robotic technologies in various domains, including transportation [29-31], health care [32,33], education [34], and defense [35,36]. In addition, researchers have developed various methods to enhance people's (proper) trust in automation or autonomy [37,38], including the use of various graphical representations [37,39-42].

For example, a military perimeter defense experiment conducted by Mercado et al [43] aimed to investigate the role of intelligent agent transparency on operator trust. Participants were tasked with selecting optimal routes for unmanned vehicles, assisted by an artificial intelligence (AI) agent. The AI agent operated at three levels of transparency: (1) basic details only; (2) basic details supplemented with reasoning and rationale; and (3) comprehensive information, including basic details, reasoning, rationale, and uncertainty indication, in a text description. They observed a positive correlation between transparency levels and participant trust. They concluded that providing operators with the agent's reasoning process and uncertainty metrics fostered a deeper understanding of the system's capabilities, thereby enhancing trust and increasing usability [43]. This finding emphasizes the importance of transparent AI systems in supporting effective human-machine collaboration.

Another study investigated the impact of visual explanations on human trust in machine learning systems [40]. Participants performed leaf classifying tasks with or without visual explanations. The leaf examples were presented in 2 formats: images and feature charts. Results revealed that providing visual explanations enhanced trust and confidence in participants' decision-making. Interestingly, the feature charts were designed with intentional omissions of detailed explanations of features to prevent information overload. However, this simplification led participants to struggle to interpret the charts, and expert users expressed a need for more comprehensive feature descriptions to inform their decisions [40]. This insight reveals the importance of integrating visual explanations with a thoughtfully managed information load for appropriate human trust.

Signal detection theory (SDT) is commonly used to study trust in automation by modeling the reliability of automated systems used by human operators. SDT evaluates how AI systems distinguish signals from noise, categorizing the state of the world as either "signal present" or "signal absent." Based on SDT categorization, AI performance results in 4 outcomes: hit (error flagged correctly), miss (error not flagged), false alarm (FA; no error, but flagged incorrectly), and correct rejection (CR; no error and no flag). Correct identifications (hit and CR) increase trust, while incorrect ones (FA and miss) decrease trust [44,45]. Research indicates that FAs typically reduce trust less than misses, prompting designers to design more liberal systems (ie, more willing to flag an error) with higher rates of FAs to ensure potential issues are flagged [46-48]. In the context of pill dispensing, FAs may lead to minor disruptions, while misses could lead to dispensing errors, indicating that a more liberal AI system prioritizing safety by minimizing misses is beneficial.

Objectives

This study, therefore, aimed to explore pharmacists' trust in automated pill verification technology, which is designed to assist with the critical task of medication dispensing. Specifically, we aimed to study the role of presenting AI uncertainty information on pharmacists' trust in the system. The primary hypothesis was as follows:

- H1: Presenting AI uncertainty information of predicted probability in a visualization format will increase AI transparency, leading to enhanced pharmacists' trust in pill verification technology.

Beyond this primary focus, we also explored how pharmacists' trust behavior varied across different AI performance patterns categorized by SDT. Drawing from these arguments, we derived the following hypotheses:

- H2: Misses would result in a more significant decline in trust compared to FAs.
- H3: Furthermore, given the differing consequences associated with the 4 SDT patterns, we speculate that presenting AI uncertainty information might have varying effects depending on the specific type of patterns.

Methods

Ethical Considerations

This research was exempt from institutional review board oversight by the University of Michigan (HUM#00213493). Before participating, participants signed an electronic informed consent form, and all data were collected anonymously. Participants received US $150 upon completion of the study.

Recruitment and Participants

Recruitment emails were dispatched to pharmacists through the Minnesota Pharmacy Practice-Based Research Network and the University of Michigan College of Pharmacy Pharmacist Preceptor Network. To meet the inclusion criteria, pharmacists were required to (1) be licensed pharmacists in the United States, (2) be aged at least 18 years, and (3) have access to a laptop or desktop computer with a webcam. Pharmacists who (1) require assistive technology to use the computer; (2) wear eyeglasses with more than one power; (3) have uncorrected cataracts, intraocular implants, glaucoma, or permanently dilated pupils; and (4) have eye movement or alignment abnormalities (eg, lazy eye, strabismus, nystagmus) were excluded from participation in the study (Figure 1). A total number of 30 licensed pharmacists in the United States completed the study. Table 1 shows the demographic information.

Figure 1. Participant recruitment timeline.

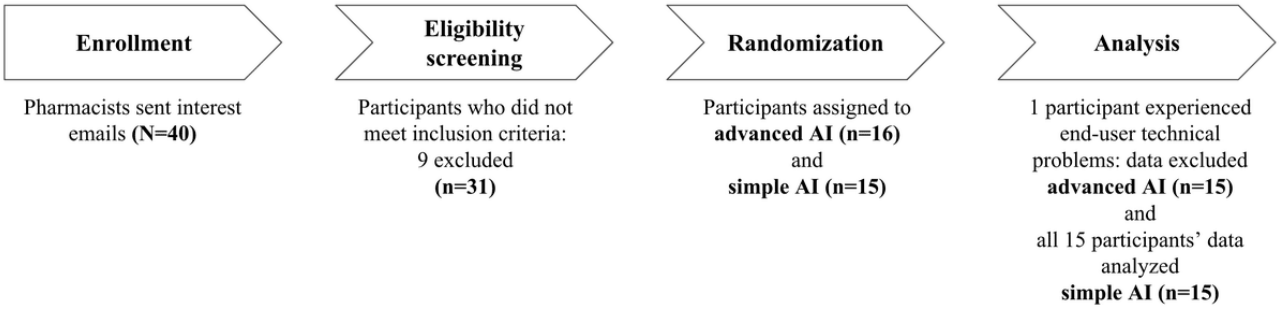


Table 1. Participant demographic information (n=30).

Characteristic	Value
Age (years), mean (SD)	39.40 (11.23)
Sex, n (%)	
Female	17 (57)
Male	13 (43)
Practice setting, n (%)	
Community pharmacy	15 (50)
Hospital pharmacy	6 (20)
Specialty pharmacy	1 (3)
Grocery store or mass merchandise pharmacy	1 (3)
Other	7 (23)
Years worked, n (%)	
1-5	7 (23)
6-10	7 (23)
11-20	10 (33)
21 or more	6 (20)

AI Model

The AI model used in this study is a Bayesian neural network that predicts the National Drug Code (NDC), a unique identifier assigned by the Food and Drug Administration to catalog drug products in the United States [49,50]. The Bayesian neural network used a technique known as random dropout [51], applied to a ResNet-34 convolutional neural network architecture [52] to estimate the probability associated with each NDC (ie, each class). The model produced 50 different predictions, where each prediction is a probability vector that quantifies the probabilities that an input belongs to each of the NDCs. The predicted NDC was then attained by finding the highest average probability derived from the 50 predictions.

To train the AI model, we acquired a dataset of 432,974 images from a mail-order pharmacy in the United States. This pharmacy uses a robotic system that counts pills, fills and labels the bottle, captures the images of the contents, and seals the bottle with a cap. The image dataset consists of 1 year's worth of robot-captured images of oral medications, such as tablets and capsules, inside prescription bottles filled by the robotic system. Each image in the dataset is associated with an NDC label and various attributes of color, shape, size, manufacturer, tablet scoring, and imprint. The number of images available for each NDC varied, ranging from 3 to 12,105, with a median of 540 (IQR 257-1291). The medications featured in these datasets were classified into 12 different colors and 7 distinct shapes. The detailed classification is shown in Table 2.

Table 2. Percentage of medication characteristics featured in the training dataset (N=260,119).

Characteristics	Dataset, n (%)
Colors	
White	109,487 (42.1)
Yellow	32,041 (12.3)
Pink	23,585 (9.1)
Orange	18,541 (7.1)
Multicolor	15,289 (5.9)
Green	13,644 (5.2)
Red	13,474 (5.2)
Blue	12,452 (4.8)
Brown	9792 (3.8)
Purple	8184 (3.1)
Turquoise	1858 (0.7)
Gray	1772 (0.7)
Shapes	
Round	128,947 (49.6)
Oval	86,844 (33.4)
Capsule	42,040 (16.2)
Hexagon (6-sided)	1150 (0.4)
Triangle	738 (0.3)
Trapezoid	280 (0.1)
Pentagon (5-sided)	120 (0)

Experimental Testbed and Stimuli

In the experiment, participants performed a pill verification task with the help of an imperfect AI aid that recommends whether to accept or reject a filled medication. The participant's task was to verify whether the filled medication matched the reference image. If the reference image and filled medication did not match, the correct action was to click "reject." If the reference image and filled medication matched, the correct action was to click "accept."

The user interface was designed following pharmacists' feedback from a focus group study conducted by the research team [50]. The interface displayed an image of a filled medication, a reference image, prescription information, and AI aids. There were 2 types of AI aids powered by the same AI model: one aid augmented with an uncertainty plot indicating the degree of certainty (or uncertainty) of the AI recommendation, and the other aid without this feature. Both AI aids recommend the action the pharmacist should take, using 4 checkboxes. A recommendation to accept was indicated when all four checkboxes turned green (Figure 2); otherwise, the recommendation was to reject. For the AI aid with the uncertainty plot, an additional histogram was integrated (Figure 3). The histogram displayed the distribution of the 50 probabilities for the predicted NDC, generated by the 50 predictions. The purpose of the histogram was to provide a visual representation of the certainty or uncertainty level associated with the AI's NDC prediction.

With the help of an AI aid, participants performed a block of 100 pill verification trials. The experimental stimuli for the 100 trials, including the reference NDC and the filled medication, were carefully selected from the dataset of 432,974 images. The selection process ensured a broad representation of colors and shapes (Table 3), while blurry images were excluded to maintain clarity. To minimize learning effects, each reference NDC was intentionally shown no more than twice throughout the experiment.

Figure 2. Interface for the AI aid without the uncertainty plot. Checkboxes indicate AI's recommendation. When all 4 checkboxes are green, the AI advises to accept; otherwise, it advises to reject. AI: artificial intelligence; NDC: National Drug Code.

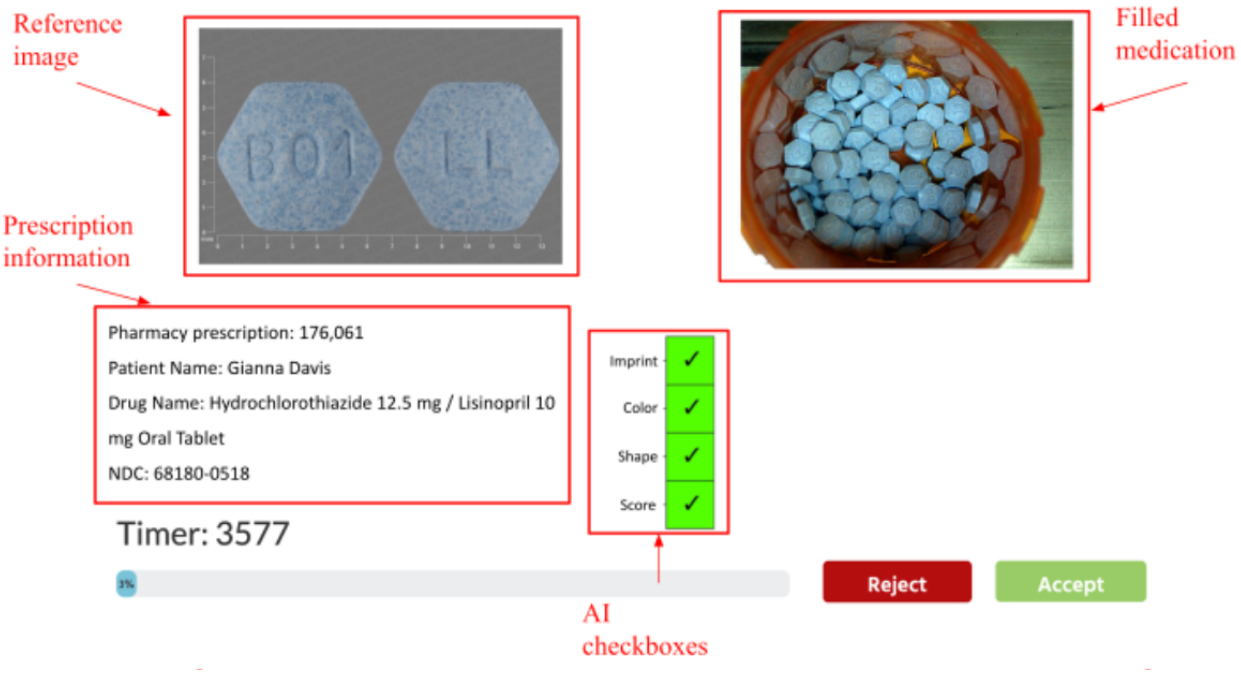


Figure 3. Interface for the AI aid with the uncertainty plot. In addition to Figure 2, the histogram shows the distribution of the 50 predicted probabilities associated with the predicted NDC. AI: artificial intelligence; NDC: National Drug Code.

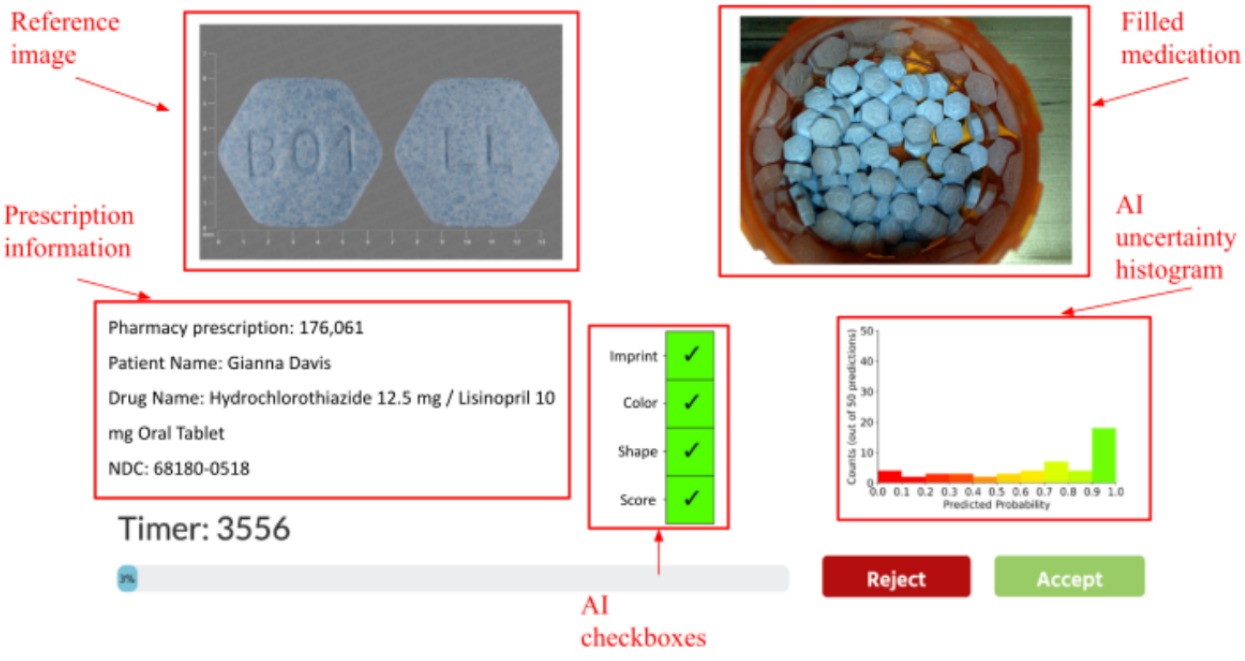
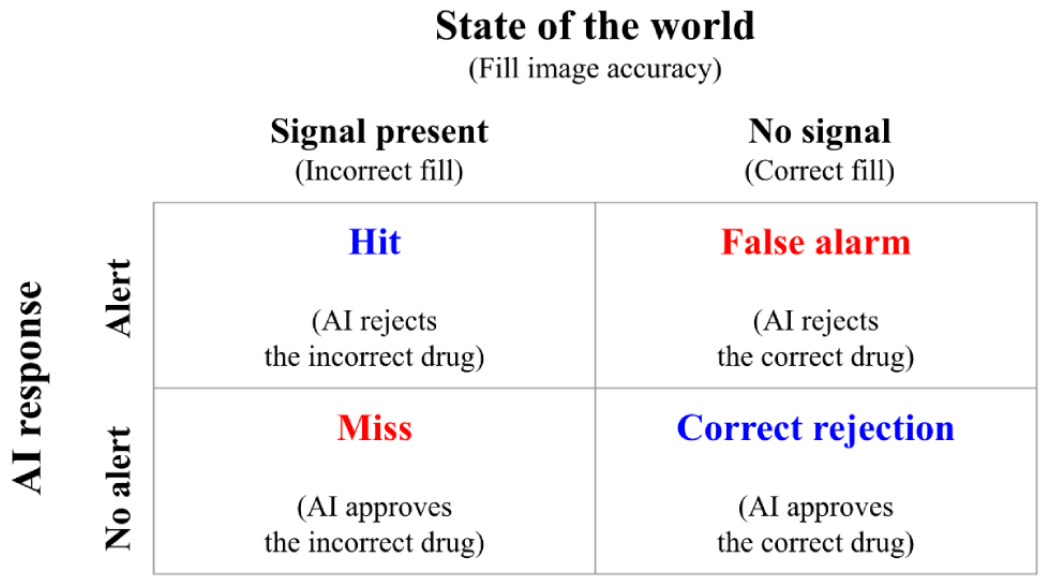


Table 3. Percentage of medication characteristics featured in artificial intelligence–aided trials (N=100) for reference images and filled images.

Characteristics	Reference, n (%)	Filled, n (%)
Color		
White	35 (35)	37 (37)
Yellow	12 (12)	12 (12)
Pink	8 (8)	8 (8)
Orange	8 (8)	8 (8)
Multicolor	3 (3)	3 (3)
Green	6 (6)	6 (6)
Red	3 (3)	2 (2)
Blue	10 (10)	10 (10)
Brown	10 (10)	10 (10)
Purple	4 (4)	3 (3)
Turquoise	1 (1)	1 (1)
Gray	1 (1)	0 (0)
Shape		
Round	54 (54)	51 (51)
Oval	24 (24)	28 (28)
Capsule	19 (19)	19 (19)
Hexagon (6-sided)	2 (2)	1 (1)
Triangle	0 (0)	0 (0)
Trapezoid	1 (1)	1 (1)
Pentagon (5-sided)	0 (0)	0 (0)

Furthermore, the AI aid was not perfect for the 100 trials, that is, it occasionally offered incorrect recommendations. Based on SDT, we mapped out the relationship between the AI's recommendation and the true state of the world [53]. In the context of this experiment, a signal in the world was an incorrectly filled medication. The extent to which the AI recommended rejecting an incorrectly filled medication reflects its ability to detect the signal. The combination of the state of the world and the AI's recommendation resulted in four potential outcomes: (1) the AI rejects the incorrect drug (hit), (2) the AI approves the incorrect drug (miss), (3) the AI rejects the correct drug (FAs), and (4) the AI approves the correct drug (CRs), as shown in Figure 4.

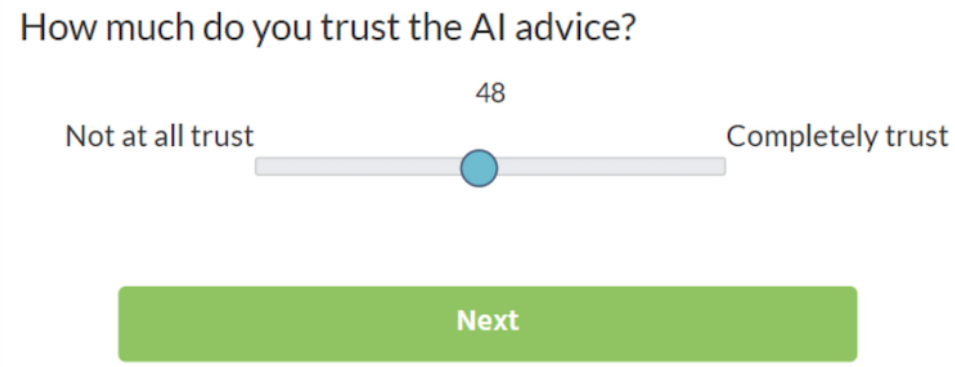
Figure 4. Four potential AI performance patterns according to signal detection theory. AI: artificial intelligence.



Benchmarking prior literature [54], the base rate was set to be 24%, that is, 24% of the trials contained wrongly filled medication. The AI accuracy was set as 82% to ensure that the AI would be perceived as useful while providing sufficient misses and false alarms [55]. By combining the filled image accuracy and AI accuracy, there were 60 cases of the AI approving the correct drug, 22 cases of the AI rejecting the incorrect drug, 2 cases of the AI approving the incorrect drug, and 16 cases of the AI rejecting the correct drug.

After each trial, participants received performance feedback indicating the correctness of their decision to accept or reject, as well as whether the prescription bottle was correctly or incorrectly filled (ie, "Your decision was correct. The medication was correctly filled"). Following this step, participants reported their trust in the recognition AI on a visual analog scale, with the leftmost point labeled "Not at all trust" and the rightmost point labeled "Completely trust" (Figure 5) [44,56,57].

Figure 5. Participants reported their trust in the recognition AI on a visual analog scale. AI: artificial intelligence.



Experimental Design

The experiment used a mixed subjects design. The between-subjects factor was the type of AI aid, distinguished by the presence or absence of an uncertainty plot. The within-subjects factor was the four potential outcomes: (1) the AI rejects the incorrect drug, (2) the AI rejects the correct drug, (3) the AI approves the incorrect drug, and (4) the AI approves the correct drug (Figure 4).

Half of the participants used the AI aid without the uncertainty plot, and the other half used the AI aid with the uncertainty plot. Each participant completed 2 blocks of 100 trials each. One block involved using the AI aid (either with or without the uncertainty plot), and the other block required participants to perform the task manually. The order of the 2 blocks was counterbalanced. Additionally, benchmarking prior literature [45], the trial sequence was fixed for the 100 trials in each block.

As this study is focused on the pharmacists' trust in AI, data from the manual task block were excluded from the analysis, concentrating the study's findings on interactions involving the AI aid.

Measures

Trust propensity

Before the experiment, we measured participants' trust propensity using the 6-item survey used by Merritt et al [58]. Trust propensity is "a stable, trait-like tendency to trust or not trust others" [59], and the propensity to trust machines reflects a person's tendency to trust machines in general rather than in a particular machine.

End Trust

End trust, *Trust*(*100*), is the participant's final trust rating after interacting with the AI help scenario.

Average Trust

Average trust denotes the mean of moment-to-moment trust ratings collected throughout the experiment.

Trust Change

After each trial *i*, participants reported their *Trust*(*i*) in the AI. We calculated a trust change as follows.

$$\text{Trust change } (i) = \text{Trust}(i) - \text{Trust}(i-1), \text{ where } i=2, 3, \ldots, 100.$$

Since the moment-to-moment trust was reported after each trial, 99 trust changes were obtained from each participant.

Experimental Procedure

The experiment was conducted remotely with interested participants who met the inclusion criteria. Each interested individual was phone screened to determine their eligibility. Before each experiment, participants had a brief web-based meeting with a member of the study team to ensure that the physical environment, including lighting conditions, was suitable for the experiment. Subsequently, the pharmacists were directed to Labvanced's website on their computer and presented with a 15-minute video tutorial that explained how to perform a simulated medication verification task using the testbed interface. Pharmacists were informed that the objective of the task was to determine whether an image of a filled medication

bottle matched a known reference image. The tutorial also explained the 2 AI aids.

Before engaging in the verification task, participants were directed to complete a demographics survey and a trust propensity survey [58]. Additionally, they went through a set of calibration procedures for the eye-tracking software. Participants then completed a block of 100 verification trials using an AI aid—either with or without the uncertainty plot—and another block of 100 verification trials manually, conducted in a counterbalanced order. Upon completion of the 200 trials, participants filled out a postexperimental survey and answered nonmandatory free-response feedback questions. All participants finished each block within the time limit of 60 minutes.

After completing all the tasks and surveys described earlier, participants were invited to a 30-minute debriefing session with one of the study team members. Study team members described 6 concepts of automation evaluation: observability, predictability, directing attention, exploring solution space, adaptability, and calibrated trust [60], and provided an example scenario for each concept. After each description and example, the participants were asked to provide their thoughts on how the concept relates to our system.

Statistical Analysis

Participants' trust propensity, end trust, and trust change when using both types of AI aids were analyzed. First, we conducted a descriptive analysis of the participants' trust propensity. Then, to test our directional hypothesis that AI aids with uncertainty will result in higher end and average trust, we conducted a 1-tailed *t* test. Finally, we analyzed how trust increased and decreased after participants experienced each of the 4 AI performance patterns using mixed-linear models with random intercept. Regression 2-tailed *t* tests were conducted to compare the means between the 2 AI aids for each AI performance pattern. The Kenward-Roger method was used to estimate degrees of freedom. Mixed model (generalized linear models) tests were conducted using the R (version 4.2.2; R Foundation for Statistical Computing) *lme4* package [61]. All statistical significance was determined at the α=.05 level and analyses were carried out using R statistical software [62].

Results

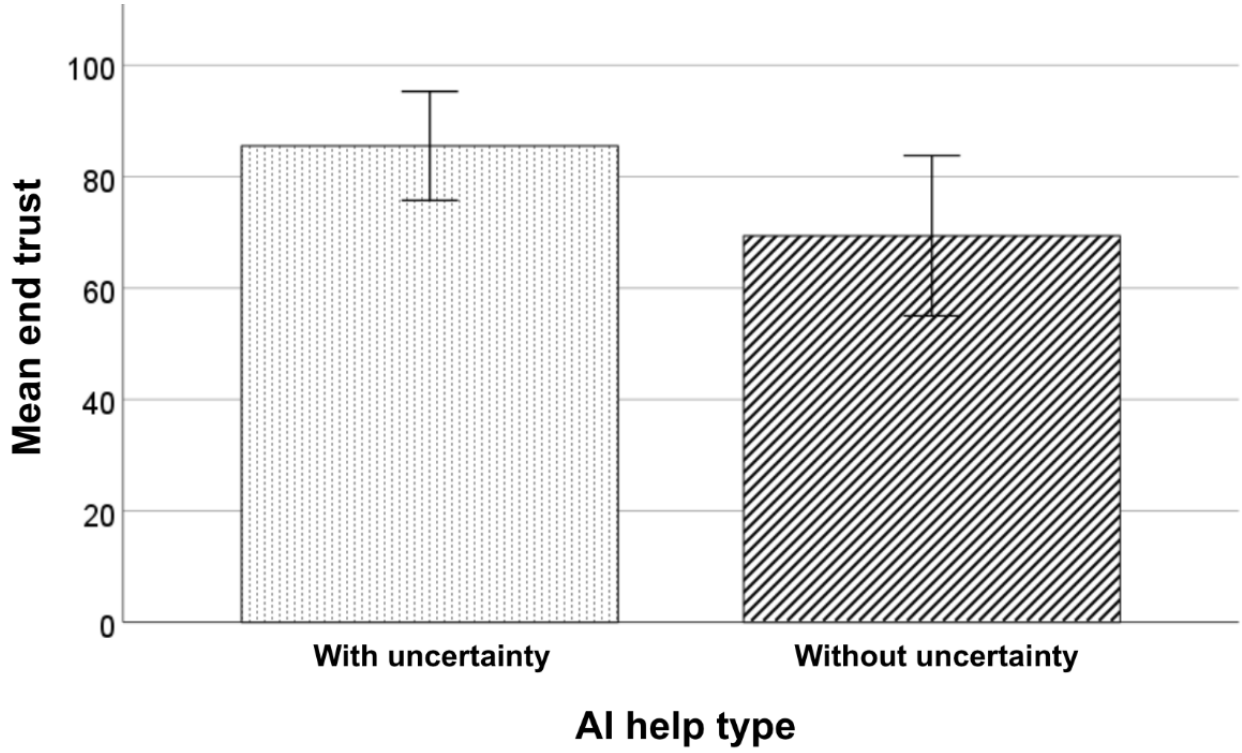
Trust Propensity

Participants had an average trust propensity score of 72 (SD 18.08) out of 100, indicating the participants generally had a positive attitude toward trusting automated technologies. There was no significant difference between the 2 AI aids (t_{28}=–0.854; *P*=.20; Cohen *d*=.312).

End Trust Toward AI Aid

The 1-tailed *t* test indicated that participants trusted the AI aid with the uncertainty plot significantly more than the AI aid without the plot at the end of the experiment (t_{28}=–1.854; *P*=.04; Cohen *d*=–.677; Figure 6).

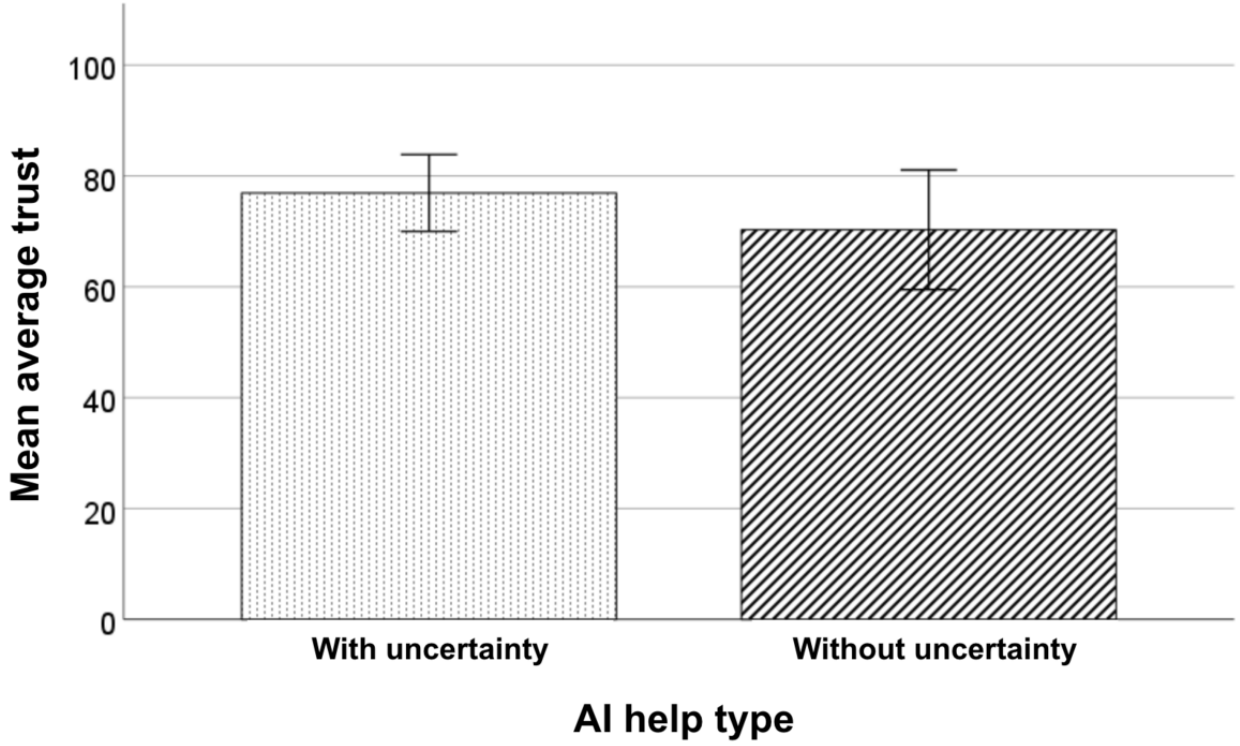
Figure 6. Mean end trust by AI aid help type (with or without the uncertainty plot). The error bars represent a 95% CI. AI: artificial intelligence.



Average Trust Toward AI Aid

Participants showed a slightly higher average trust in the AI aid in the with-uncertainty condition (mean 76.92, SD 13.42) than in the without-uncertainty condition (mean 70.29, SD 20.88; Figure 7). However, the difference did not reach statistical significance (t_{28}=–1.036; *P*=.16; Cohen *d*=–.378).

Figure 7. Mean average trust by AI aid help type (with or without the uncertainty plot). The error bars represent a 95% CI. AI: artificial intelligence.



Trust Change

Figure 8 shows the trust change after participants experienced the 4 AI performance patterns. When the AI approved the correct drug, there was a significantly greater trust increment when participants used the AI aid with the uncertainty plot compared to the one without ($t_{78.98}$=3.93; *P*<.001; Cohen *d*=.214); When the AI approved the incorrect drug, there was a significantly greater trust decrement using the AI aid with the uncertainty plot ($t_{2939.72}$=–4.78; *P*<.001; Cohen *d*=.712). Interestingly, when the AI rejected the incorrect drug, we observed a decrement of trust for participants using the AI aid with the uncertainty plot ($t_{509.77}$=–3.96; *P*<.001; Cohen *d*=.312). When the AI rejected the correct drug, both AI help types showed a decrease in trust and there was no statistical difference between them ($t_{856.57}$=–0.68; *P*=.49; Cohen *d*=.045). Overall, participants using the AI aid with the uncertainty plot displayed a large magnitude of trust adjustment. In addition, we observed a significant "negativity bias" in that the magnitude of trust change when the AI made an error (ie, the AI approves the incorrect drug or the AI rejects the correct drug) was significantly larger than the magnitude of trust adjustment when the AI provided correct recommendations (generalized linear model test; *z*=–11.30; *P*<.001).

To examine variation in pharmacists' trust behavior across different AI performance patterns categorized by SDT, we initially combined the data from both the with and without uncertainty help scenarios. Trust decreased significantly more when the AI approved the incorrect drug compared to when the AI rejected the correct drug (t_{509}=–4.687; *P*<.001; Cohen *d*=.475). This trend was observed in the with-uncertainty AI help scenario (t_{254}=–4.91; *P*<.001; Cohen *d*=.758). However, in the without-uncertainty AI help scenario, although there was a greater trust decrease in trust when the AI approved the incorrect drug than when the AI rejected the correct drug, the difference was not statistically significant (t_{254}=–1.14; *P*=.255; Cohen *d*=.014).

As we measured participants' trust toward AI continuously, we calculated the autocorrelation between the trust ratings. Autocorrelation measures the relationship between a variable's current value and its past values in a time series. Figure 9 shows the mean autocorrelation as a function of time separation between the ratings. For both AI aids, the correlation decreased as the time separation increased. The AI aid with the uncertainty plot had a lower mean autocorrelation compared to the aid without the uncertainty plot (Figure 9).

Figure 8. Trust change for different AI performance patterns. Red shades represent negative trust change, and blue shades represent positive trust change.

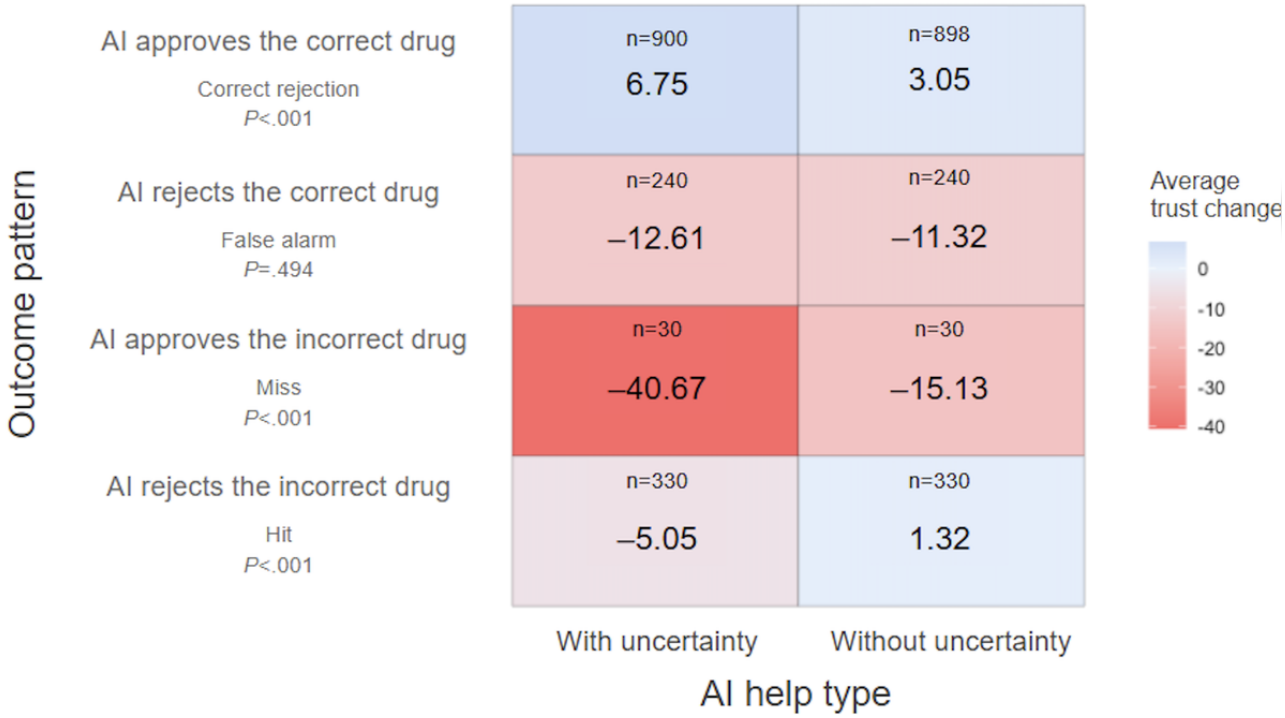
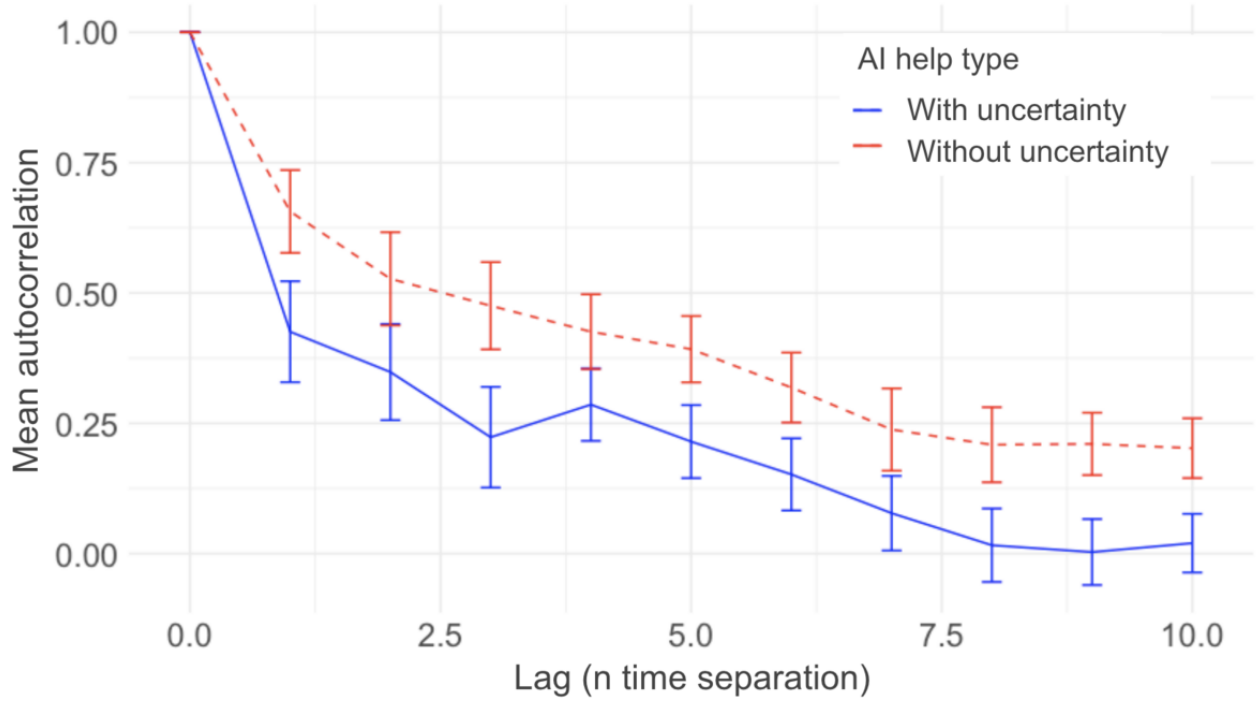


Figure 9. Autocorrelation of trust as a function of time separation. The blue solid line represents AI aid with the uncertainty plot, and the red dashed line represents the AI aid without the uncertainty plot. The error bars represent 2 SEs. AI: artificial intelligence.



Discussion

Principal Findings

This study aimed to investigate pharmacists' trust in automated pill verification technology and how the presentation of AI uncertainty information would influence them. Overall, the findings revealed that pharmacists have a favorable disposition toward trusting automation, and including the AI's uncertainty information increases pharmacists' trust in the AI recommendation *(*Figure 6*)*.

Comparison With Prior Work

The propensity to trust automation refers to an individual's general inclination to trust automated or autonomous systems, shaped by their past experiences and future expectations [58,59]. Research has shown that levels of trust propensity vary among

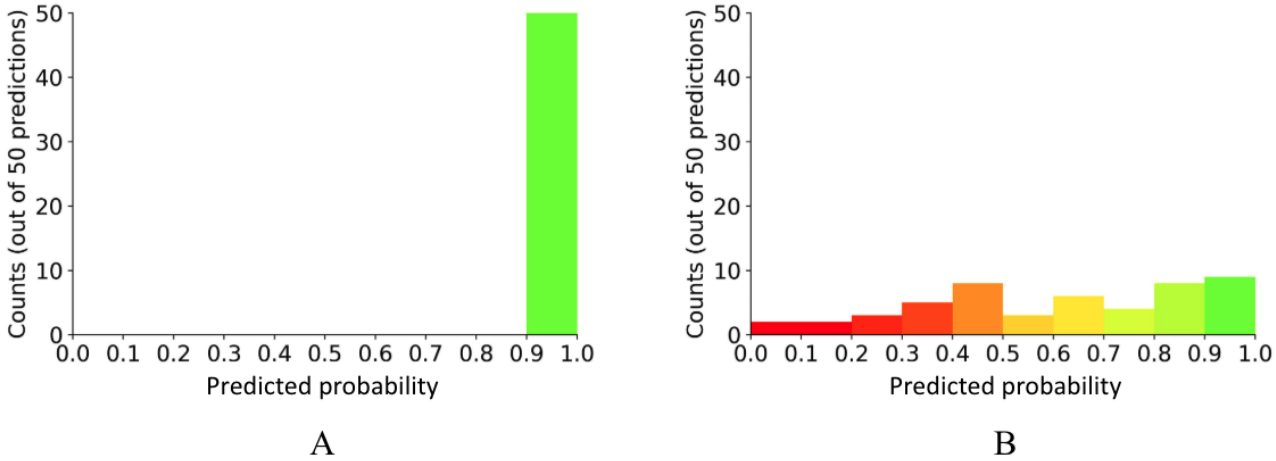
individuals. For example, an early study by Merritt et al [58], which included 69 college students (average age of 25 years), found an average trust propensity of 3.56 (SD 0.6) on a 7-point Likert scale [59]. More recently, Montag et al [63] surveyed 289 participants aged between 18 and 70 years and reported their propensity to trust automation to be 4.98 (SD 1.06) after converting to a 7-point Likert scale. Similarly, Miller et al [64] reported a trust propensity score of 4.97 (SD 1.21) from a smaller cohort of 28 participants aged between 18 to 60 years. Another investigation by Yang et al [65] with 75 adults (mean age 23.0) split into 3 groups reported trust propensity scores of 72.6 (SD 14.8), 69.4 (SD 10.4), and 69.4 (SD 14.4), equivalent to average scores of 5.08 (SD 1.89), 4.86 (SD 1.62) and 4.86 (SD 1.86) on a 7-point Likert scale.

In line with these findings [63-66], our study revealed that pharmacists generally have a favorable disposition toward trusting automation, with an average rating of 72 (SD 18.08) on a 100-point scale, or 5.03 (SD 2.09) on a 7-point Likert scale. This positive attitude may be attributed to the frequent use of automated technologies, such as barcode scanners and pill counters in their daily work [6,7]. Additionally, as expected, no significant difference was observed between the groups using different AI aids because the participants were randomly assigned to use either of the AI aids.

Examining pharmacists' end trust, our findings reveal that the AI aid with the uncertainty plot significantly enhanced the end trust scores. We attribute this enhancement to the increased transparency achieved through the presentation of a histogram showing the distribution of 50 predicted probabilities. While AI advancements promise to improve human performance, a prevailing issue is the perception of AI as a "black box." This lack of transparency contributes to a lack of trust in AI and can undermine team performance [66-68]. The higher end trust observed in participants using the AI aid with the uncertainty plot indicates that making the AI more transparent by revealing its decision-making process can foster a higher level of trust in automation. Participants 14 and 24 captured this sentiment well stating: "As soon as the uncertainty plot became red and yellow, I slowed down, which could be helpful because sometimes slowing down when there is uncertainty, just knowing there's uncertainty, is enough" (P14) and "if the uncertainty plot was all green bar and AI thought it was doing a 100% accurate job then it was easier to make my decision" (P24).

Regarding the dynamics of trust, that is, moment-to-moment trust change, when AI approved the correctly filled bottle, we noted trust increments for both AI aids. Furthermore, the inclusion of uncertainty information led to a larger increment in trust compared to when such information was absent. When the AI mistakenly approved the incorrect drug, we observed a significant trust decrement for both AI aids, potentially attributed to the adverse outcome associated with the wrong medication. Furthermore, the trust decrement was significantly larger when the uncertainty information was shown. This pronounced trust decrement could have resulted directly from the distribution of the histogram: participants were shown a histogram indicating a high level of certainty (Figure 10A). Therefore, participants may have perceived the error made by the AI aid as a "confident" error and therefore reduced their trust even more. Studies examining likelihood alarms reported that highly likely alarms (ie, "confident" alarms) engender a greater decline in momentary trust upon automation failures [57,69].

Figure 10. Uncertainty plots with (A) narrower and (B) wider IQRs. IQR is a measure of statistical dispersion, or how spread out the data points are.

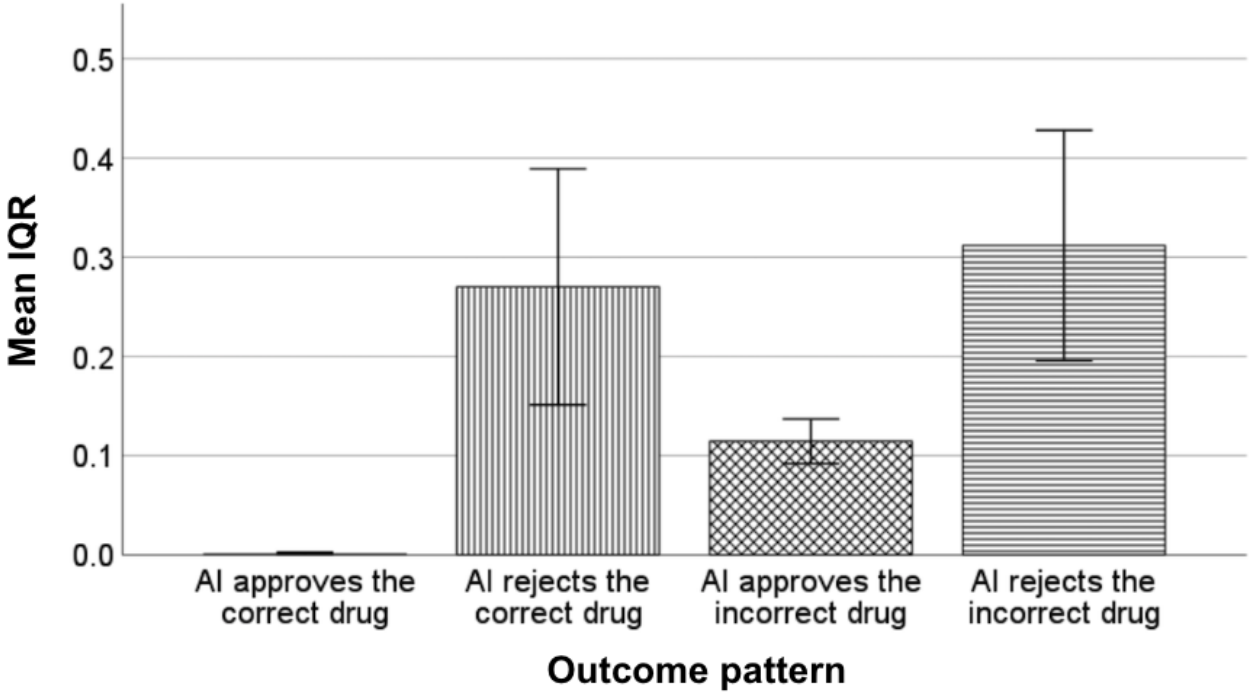


Our study also offers additional validation of prior findings regarding the SDT modeling of trust [47,48]. When the AI mistakenly approved the incorrect drug (miss), a greater trust decrease appeared compared to when the AI rejected the correctly filled bottle (FA). However, the difference was statistically significant only in the with-uncertainty AI help type. This finding may provide further evidence that "confident" AI errors lead to a greater trust decrement [57,69].

Intriguingly, when the AI rejected an incorrectly filled bottle, the absence of uncertainty information resulted in an increase in trust, whereas the presence of such information led to a decrease in trust. Such contrasting results could have stemmed from the uncertainty plots influencing the participants' decision-making process. When the AI approved the correct drug, all uncertainty plots presented to participants showed a consistent solid green bar (Figure 10A). However, when the AI rejected the incorrect drug, the IQR of the uncertainty plot was broader, indicating the lack of certainty (Figure 11). A total of 16 (73%) out of 22 uncertainty plots displayed a wider spread with mixed color bars (Figure 10B). This ambiguity might

unintentionally cause the human participants to doubt the capability of the AI aid with the uncertainty plot, resulting in a decrease in their trust, as evidenced by P18's statement: "When the checkboxes were not all green and the histogram had a bunch of colors (variability), I was even less trusting of the AI tool." This perspective is also supported by P25, who noted: "When the uncertainty plot had some red, it made me double check and decreased my confidence for sure."

Figure 11. Mean IQR by outcome pattern. The error bars represent a 95% CI. AI: artificial intelligence.



Finally, when AI rejected the correctly filled bottle, trust decrements occurred, with no significant differences between the 2 AI aids. If this circumstance happened in the real world, the pharmacists would reinspect the filled prescription. It will likely lead to an increased workload, fatigue, and stress, which could potentially lead to a lower quality of work and a higher frequency of errors [70,71]. However, P21 offered a contrasting viewpoint that more flagging would be better than AI approving the incorrect, highlighting: "I didn't lose as much trust when AI rejected the correct drug. I feel like AI should be there as a cautionary tool."

The observed trends in the moment-to-moment dynamics of trust indicate a greater degree of trust adjustment when participants were assisted by the AI with the uncertainty plot. This observation is further confirmed by the autocorrelation analysis. Specifically, the trust autocorrelation plot (Figure 9) reveals a lower autocorrelation between trust ratings when the uncertainty information was presented. This suggests that current trust levels were less influenced by past trust levels, implying more significant changes in trust from moment to moment. Pharmacists relied less on previous trials and more on the information presented in the present trial, highlighting the advantages of a more transparent display.

In addition, for both AI aids, participants displayed a larger trust decrement due to incorrect automation predictions. Even though these observations may seem alarming initially, they align with the prior literature addressing negativity bias. The study suggests that failure in automation typically has a more significant negative impact on trust than a positive impact from successful automation [56,65].

Limitations and Future Directions

We acknowledge several limitations of this study and propose directions for future research. First, as a pioneering investigation in this domain, we did not strictly control the interquartile range of the uncertainty plot (Figure 11). Future investigation should systematically examine the effects of presenting different distributions within the uncertainty plot on pharmacists' trust. Exploring similar variations in the IQR among different outcome patterns could provide a deeper understanding of how outcome patterns could impact trust while avoiding potential confounding factors.

Second, the uncertainty plot used in this study displayed the distribution of only 1 (2%) out of 50 probabilities for the predicted NDC. Future research should consider incorporating additional contextual information to enhance the interpretability of these distributions. A notable challenge identified was that users unfamiliar with statistical representation found the uncertainty plot difficult to understand. P11 suggested simplifying the uncertainty presentation because "it would be a little bit too much for some people not as comfortable with statistics or technology." Nonetheless, there is potential for pharmacists to become more comfortable with these plots with prolonged use and training, as evidenced by P30's remark: "I started skeptical because it's something I'm not familiar with, but as I got more examples of it, my trust built up quickly." Consequently, researchers should continue to develop alternative visualization techniques that provide a more comprehensive and intuitive representation of the AI's uncertainty, while maintaining a low complexity to accommodate users with varying degrees of statistical proficiency.

Third, this exploratory study was limited by a small sample size, which may have impacted the statistical power of the analyses on trust propensity, end trust, and average trust. This limitation likely contributed to the lack of significance observed in some results. However, despite being underpowered, our analysis still revealed a significant difference in end trust, indicating a large effect size. Regarding trust change, to maintain the AI's perceived usefulness while still providing enough examples of both misses and FAs, we incorporated only 2 trials of the AI approving the incorrect drug (miss) and 16 trials of the AI rejecting the correct drug (FA), setting the accuracy at 82%. The statistical power to detect significant differences may have been compromised by the small sample sizes. Future research should aim to include larger participant pools to enhance the generalizability and robustness of the findings.

Finally, this study only focused on pharmacists' trust and trust change and did not include the analysis of accuracy and reaction time. Even though focusing on trust alone is an accepted practice [53], a more comprehensive analysis linking performance with trust would likely reveal the relationship between performance and trust calibration.

Conclusions

Dispensing errors are significant contributors to adverse drug events, which lead to considerable health care expenses and harm to patients. Despite progress made in developing automated technologies to aid pill verification, pharmacists' trust in these systems has not been thoroughly investigated. Our research represents an initial exploration into pharmacists' trust in automated pill verification technology, marking a significant step in understanding the integration of such systems into health care settings.

Our findings reveal that pharmacists have a favorable disposition toward trusting automation, which can likely be attributed to their frequent use of automated technologies in their daily work. Moreover, providing uncertainty information about the AI's recommendation significantly boosts pharmacists' trust in the AI aid, highlighting the importance of transparency in AI development. The dynamics of trust vary depending on the AI's performance. Pharmacists using the AI aid with the uncertainty plot had a significantly larger trust increment when the AI approved the correct drug and a significantly larger trust decrement when the AI approved the incorrect drug. Intriguingly, the absence of the uncertainty plot led to an increase in trust when the AI correctly rejected an incorrect drug, whereas the presence of the plot resulted in a decrease in trust under the same circumstances. In addition, a pronounced "negativity bias" was observed, where the degree of trust reduction when the AI made an error exceeded the trust gain when the AI made a correct decision.

Acknowledgments

This research study was supported by the National Library of Medicine of the National Institutes of Health (R01LM013624). The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

Data Availability

The deidentified data produced in this study are available upon reasonable request through a data use agreement with the University of Michigan.

Authors' Contributions

The authors confirm their contribution to the article as follows: study conception and design: JYK conducted subject testing, conducted statistical analysis, interpreted the results, and wrote the manuscript, VDM conducted statistical analysis and interpreted the results, BR conducted subject testing, QC interpreted the results, provided data for Tables 1, 2, and 3, YZ conducted subject testing, JDL interpreted the results and provided suggestion for Figure 9, RAK interpreted the results, CL interpreted the results, and XJY conducted statistical analysis, interpreted the results, and wrote the manuscript; All authors designed and conceptualized the study, reviewed the results, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Reiner G, Pierce SL, Flynn J. Wrong drug and wrong dose dispensing errors identified in pharmacist professional liability claims. J Am Pharm Assoc 2020;60(5):e50-e56. [doi: 10.1016/j.japh.2020.02.027] [Medline: 32217085]
2. Chui MA, Look KA, Mott DA. The association of subjective workload dimensions on quality of care and pharmacist quality of work life. Res Social Adm Pharm 2014;10(2):328-340 [FREE Full text] [doi: 10.1016/j.sapharm.2013.05.007] [Medline: 23791360]
3. Chui MA, Mott DA. Community pharmacists' subjective workload and perceived task performance: a human factors approach. J Am Pharm Assoc 2012;52(6):e153-e160 [FREE Full text] [doi: 10.1331/JAPhA.2012.11135] [Medline: 23229977]

4. Balchaites A, McCarthy S, Fleming A. Exploring the human factors of medication errors in community pharmacy: a mixed methods study. Int J Pharm Pract 2022:i39-i40. [doi: 10.1093/ijpp/riac019.055]
5. Larios Delgado N, Usuyama N, Hall AK, Hazen RJ, Ma M, Sahu S, et al. Fast and accurate medication identification. NPJ Digit Med 2019;2(1):10 [FREE Full text] [doi: 10.1038/s41746-019-0086-0] [Medline: 31304359]
6. Bates D. Using information technology to reduce rates of medication errors in hospitals. BMJ 2000;320(7237):788-791 [FREE Full text] [doi: 10.1136/bmj.320.7237.788] [Medline: 10720369]
7. Poon EG, Cina JL, Churchill W, Patel N, Featherstone E, Rothschild JM, et al. Medication dispensing errors and potential adverse drug events before and after implementing bar code technology in the pharmacy. Ann Intern Med 2006;145(6):426-434. [doi: 10.7326/0003-4819-145-6-200609190-00006] [Medline: 16983130]
8. Poon EG, Keohane CA, Yoon CS, Ditmore M, Bane A, Levtzion-Korach O, et al. Effect of bar-code technology on the safety of medication administration. N Engl J Med 2010;362(18):1698-1707. [doi: 10.1056/NEJMsa0907115] [Medline: 20445181]
9. Paoletti R, Suess T, Lesko M, Feroli AA, Kennel JA, Mahler JM, et al. Using bar-code technology and medication observation methodology for safer medication administration. Am J Health Syst Pharm 2007;64(5):536-543. [doi: 10.2146/ajhp060140] [Medline: 17322168]
10. Young D. Veterans Affairs bar-code-scanning system reduces medication errors. Am J Health Syst Pharm 2002;59(7):591-592. [doi: 10.1093/ajhp/59.7.591] [Medline: 11944599]
11. Debono DS, Greenfield D, Travaglia JF, Long JC, Black D, Johnson J, et al. Nurses' workarounds in acute healthcare settings: a scoping review. BMC Health Serv Res 2013;13:175 [FREE Full text] [doi: 10.1186/1472-6963-13-175] [Medline: 23663305]
12. Koppel R, Wetterneck T, Telles JL, Karsh B. Workarounds to barcode medication administration systems: their occurrences, causes, and threats to patient safety. J Am Med Inform Assoc 2008;15(4):408-423 [FREE Full text] [doi: 10.1197/jamia.M2616] [Medline: 18436903]
13. Oldland AR, Golightly LK, May SK, Barber GR, Stolpman NM. Electronic inventory systems and barcode technology: impact on pharmacy technical accuracy and error liability. Hosp Pharm 2015;50(1):34-41 [FREE Full text] [doi: 10.1310/hpj5001-34] [Medline: 25684799]
14. Buchanan TL, Barker KN, Gibson JT, Jiang BC, Pearson RE. Illumination and errors in dispensing. Am J Hosp Pharm 1991;48(10):2137-2145. [Medline: 1781468]
15. Gidman W. Increasing community pharmacy workloads in England: causes and consequences. Int J Clin Pharm 2011;33(3):512-520. [doi: 10.1007/s11096-011-9498-x] [Medline: 21424615]
16. Lea VM, Corlett SA, Rodgers RM. Workload and its impact on community pharmacists' job satisfaction and stress: a review of the literature. Int J Pharm Pract 2012;20(4):259-271. [doi: 10.1111/j.2042-7174.2012.00192.x] [Medline: 22775522]
17. Peterson GM, Wu MSH, Bergin JK. Pharmacist's attitudes towards dispensing errors: their causes and prevention. J Clin Pharm Ther 1999;24(1):57-71. [doi: 10.1046/j.1365-2710.1999.00199.x] [Medline: 10319909]
18. Caban JJ, Rosebrock A, Yoo TS. Automatic identification of prescription drugs using shape distribution models. 2012 Presented at: 19th IEEE International Conference on Image Processing; October 3, 2012; Orlando, FL p. 1005-1008.
19. Hamilton RN. Pharmacy pill counting vision system. Google Patents. 2004. URL: https://patents.google.com/patent/US6574580B2/en [accessed 2025-01-10]
20. Popovich J, Jordan ML. Automated drug discrimination during dispensing. US Patent. 2011. URL: https://scholar.google.com/citations?view_op=view_citation&hl=en&user=YAZqaxkAAAAJ&citation_for_view=YAZqaxkAAAAJ:WF5omc3nYNoC [accessed 2025-01-13]
21. Wootton JR, Reznack VV, Hobson G. Pharmaceutical pill recognition and verification system. Google Patents. 2003. URL: https://patents.google.com/patent/US6535637B1/en [accessed 2025-01-10]
22. Yu J, Chen Z, Kamata S. Pill recognition using imprint information by two-step sampling distance sets. 2014 Presented at: 22nd International Conference on Pattern Recognition; August 24-28, 2014; Stockholm, Sweden p. 3156-3161.
23. Yu J, Chen Z, Kamata S, Yang J. Accurate system for automatic pill recognition using imprint information. IET Image Processing 2015;9(12):1039-1047. [doi: 10.1049/iet-ipr.2014.1007]
24. Lester CA, Li J, Ding Y, Rowell B, Yang, Kontar RA. Performance evaluation of a prescription medication image classification model: an observational cohort. NPJ Digit Med 2021;4(1):118 [FREE Full text] [doi: 10.1038/s41746-021-00483-8] [Medline: 34315995]
25. Wong YF, Ng HT, Leung KY, Chan KY, Chan SY, Loy CC. Development of fine-grained pill identification algorithm using deep convolutional network. J Biomed Inform 2017;74:130-136 [FREE Full text] [doi: 10.1016/j.jbi.2017.09.005] [Medline: 28923366]
26. Lee JD, See KA. Trust in automation: designing for appropriate reliance. Hum Factors 2004;46(1):50-80. [doi: 10.1518/hfes.46.1.50_30392] [Medline: 15151155]
27. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]

28. Schaefer KE, Chen JYC, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: implications for understanding autonomy in future systems. Hum Factors 2016;58(3):377-400. [doi: 10.1177/0018720816634228] [Medline: 27005902]
29. Azevedo-Sa H, Jayaraman SK, Esterwood CT, Yang XJ, Robert LP, Tilbury DM. Real-time estimation of drivers' trust in automated driving systems. Int J Soc Robot 2020;13(8):1911-1927. [doi: 10.1007/s12369-020-00694-1]
30. Lee JD, Kolodge K. Exploring trust in self-driving vehicles through text analysis. Hum Factors 2020;62(2):260-277. [doi: 10.1177/0018720819872672] [Medline: 31502885]
31. Ayoub J, Avetisian L, Yang XJ, Zhou F. Real-time trust prediction in conditionally automated driving using physiological measures. IEEE Trans Intell Transport Syst 2023;24(12):14642-14650. [doi: 10.1109/tits.2023.3295783]
32. Gombolay M, Yang XJ, Hayes B, Seo N, Liu Z, Wadhwania S, et al. Robotic assistance in the coordination of patient care. Int J Robot Res 2018;37(10):1300-1316. [doi: 10.1177/0278364918778344]
33. Wang EH, Gross CP, Tilburt JC, Yu JB, Nguyen PL, Smaldone MC, et al. Shared decision making and use of decision AIDS for localized prostate cancer : perceptions from radiation oncologists and urologists. JAMA Intern Med 2015;175(5):792-799. [doi: 10.1001/jamainternmed.2015.63] [Medline: 25751604]
34. Nazaretsky T, Ariely M, Cukurova M, Alexandron G. Teachers' trust in AI‐powered educational technology and a professional development program to improve it. Brit J Educational Tech 2022;53(4):914-931. [doi: 10.1111/bjet.13232]
35. Wang L, Jamieson GA, Hollands JG. Trust and reliance on an automated combat identification system. Hum Factors 2009;51(3):281-291. [doi: 10.1177/0018720809338842] [Medline: 19750792]
36. Du N, Huang KY, Yang XJ. Not all information is equal: effects of disclosing different types of likelihood information on trust, compliance and reliance, and task performance in human-automation teaming. Hum Factors 2020;62(6):987-1001. [doi: 10.1177/0018720819862916] [Medline: 31348863]
37. Chatzimparmpas A, Martins RM, Jusufi I, Kucher K, Rossi F, Kerren A. The state of the art in enhancing trust in machine learning models with the use of visualizations. Comput Graph Forum 2020;39(3):713-756. [doi: 10.1111/cgf.14034]
38. Rudin C, Ustun B. Optimized scoring systems: toward trust in machine learning for healthcare and criminal justice. Interfaces 2018;48(5):449-466. [doi: 10.1287/inte.2018.0957]
39. Dasgupta A, Lee J, Wilson R, Lafrance RA, Cramer N, Cook K, et al. Familiarity vs trust: a comparative study of domain scientists' trust in visual analytics and conventional analysis methods. IEEE Trans Vis Comput Graph 2017;23(1):271-280. [doi: 10.1109/TVCG.2016.2598544] [Medline: 27608465]
40. Yang F, Huang Z, Scholtz J, Arendt DL. How do visual explanations foster end users' appropriate trust in machine learning? 2020 Presented at: IUI '20: Proceedings of the 25th International Conference on Intelligent User Interfaces; March 17, 2020; Cagliari, Italy p. 189-201.
41. Sacha D, Sedlmair M, Zhang L, Lee JA, Peltonen J, Weiskopf D, et al. Human-centered machine learning through interactive visualization. 2016 Presented at: ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; April 27-29, 2016; Bruges, Belgium. [doi: 10.1016/j.neucom.2017.01.105]
42. Fisher D, Popov I, Drucker S, schraefel MC. Trust me, i'm partially right: incremental visualization lets analysts explore large datasets faster. 2012 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 5, 2012; Austin, TX p. 1673-1682.
43. Mercado JE, Rupp MA, Chen JYC, Barnes MJ, Barber D, Procci K. Intelligent agent transparency in human-agent teaming for multi-UxV management. Hum Factors 2016;58(3):401-415. [doi: 10.1177/0018720815621206] [Medline: 26867556]
44. Yang XJ, Schemanske C, Searle C. Toward quantifying trust dynamics: how people adjust their trust after moment-to-moment interaction with automation. Hum Factors 2023;65(5):862-878 [FREE Full text] [doi: 10.1177/00187208211034716] [Medline: 34459266]
45. Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. Ergonomics 1992;35(10):1243-1270. [doi: 10.1080/00140139208967392] [Medline: 1516577]
46. Swets JA. The science of choosing the right decision threshold in high-stakes diagnostics. Am Psychol 1992;47(4):522-532. [doi: 10.1037//0003-066x.47.4.522] [Medline: 1595983]
47. Davenport RB, Bustamante EA. Effects of false-alarm vs. miss-prone automation and likelihood alarm technology on trust, reliance, and compliance in a miss-prone task. Proc Hum Factors Ergon Soc Annu Meet 2010;54(19):1513-1517. [doi: 10.1177/154193121005401933]
48. Stanton NS, Ragsdale SA, Bustamante EA. The effects of system technology and probability type on trust, compliance, and reliance. HFES Ann Conf Proc 2009;53(18):1368-1372. [doi: 10.1518/107118109x12524443347238]
49. Chen Q, Al Kontar R, Nouiehed M, Yang XJ, Lester C. Rethinking cost-sensitive classification in deep learning via adversarial data augmentation. INFORMS J Data Sci 2022. [doi: 10.1287/ijds.2022.0033]
50. Zheng Y, Rowell B, Chen Q, Kim JY, Kontar RA, Yang XJ, et al. Designing human-centered AI to prevent medication dispensing errors: focus group study with pharmacists. JMIR Form Res 2023;7:e51921 [FREE Full text] [doi: 10.2196/51921] [Medline: 38145475]
51. Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. Proc Mach Learn Res 2016:1050-1059.

52. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV p. 770-778.
53. Tanner WP, Swets JA. A decision-making theory of visual detection. Psychol Rev 1954;61(6):401-409. [doi: 10.1037/h0058700]
54. Reilly S, Grasha AF, Matthews G, Schafer J. Automatic-controlled information processing and error detection in a simulated pharmacy-verification task. Percept Mot Skills 2003;97(1):151-174. [doi: 10.2466/pms.2003.97.1.151] [Medline: 14604036]
55. Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: a synthesis of the literature. Theor Issues Ergon Sci 2007;8(3):201-212. [doi: 10.1080/14639220500370105]
56. Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids. J Cogn Eng Decis Mak 2012;6(1):57-87. [doi: 10.1177/1555343411433844]
57. Yang XJ, Unhelkar VV, Li K, Shah JA. Evaluating effects of user experience and system transparency on trust in automation. 2017 Presented at: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction; March 6, 2017; Vienna, Austria p. 408-416.
58. Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: effects of implicit attitudes toward automation on trust in an automated system. Hum Factors 2013;55(3):520-534. [doi: 10.1177/0018720812465081] [Medline: 23829027]
59. Merritt SM, Ilgen DR. Not all trust is created equal: dispositional and history-based trust in human-automation interactions. Hum Factors 2008;50(2):194-210. [doi: 10.1518/001872008X288574] [Medline: 18516832]
60. McDermott P, Dominguez C, Kasdaglis N, Ryan M, Trhan I, Nelson A. Human-machine teaming systems engineering guide. Defense Technical Information Center. 2018. URL: https://apps.dtic.mil/sti/citations/AD1108020 [accessed 2025-01-10]
61. Bates DM, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Software 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
62. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2019. URL: https://www.R-project.org [accessed 2025-01-20]
63. Montag C, Kraus J, Baumann M, Rozgonjuk D. The propensity to trust in (automated) technology mediates the links between technology self-efficacy and fear and acceptance of artificial intelligence. Comput Human Behav 2023;11:100315. [doi: 10.1016/j.chbr.2023.100315]
64. Miller L, Kraus J, Babel F, Baumann M. More than a feeling-interrelation of trust layers in human-robot interaction and the role of user dispositions and state anxiety. Front Psychol 2021;12:592711 [FREE Full text] [doi: 10.3389/fpsyg.2021.592711] [Medline: 33912098]
65. Yang XJ, Guo Y, Schemanske C. From trust to trust dynamics: combining empirical and computational approaches to model and predict trust dynamics in human-autonomy interaction. In: Duffy VG, Landry SJ, Lee JD, Stanton N, editors. Human-Automation Interaction: Transportation. Cham, Switzerland: Springer; 2022:253-265.
66. de Visser EJ, Pak R, Shaw TH. From 'automation' to 'autonomy': the importance of trust repair in human-machine interaction. Ergonomics 2018;61(10):1409-1427. [doi: 10.1080/00140139.2018.1457725] [Medline: 29578376]
67. Lyons JB, Sadler GG, Koltai K, Battiste H, Ho NT, Hoffmann LC, et al. Shaping Trust Through Transparent Design: Theoretical and Experimental Guidelines. Cham, Switzerland: Springer; 2017:127-136.
68. Luo R, Du N, Yang XJ. Evaluating effects of enhanced autonomy transparency on trust, dependence, and human-autonomy team performance over time. Int J Hum Comput 2022;38(18-20):1962-1971. [doi: 10.1080/10447318.2022.2097602]
69. Sorkin RD, Kantowitz BH, Kantowitz SC. Likelihood alarm displays. Hum Factors 1988;30(4):445-459. [doi: 10.1177/001872088803000]
70. Carayon P, Gurses AP. Nursing workload and patient safety—a human factors engineering perspective. In: Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
71. Watterson TL, Look KA, Steege L, Chui MA. Operationalizing occupational fatigue in pharmacists: an exploratory factor analysis. Res Social Adm Pharm 2021;17(7):1282-1287. [doi: 10.1016/j.sapharm.2020.09.012] [Medline: 33004303]

Abbreviations

AI: artificial intelligence
CR: correct rejection
FA: false alarm
NDC: National Drug Code
SDT: signal detection theory

Edited by A Kushniruk, E Borycki; submitted 06.05.24; peer-reviewed by J Waterson, HF Neyedli; comments to author 13.09.24; revised version received 08.11.24; accepted 22.12.24; published 11.02.25.

Please cite as:
Kim JY, Marshall VD, Rowell B, Chen Q, Zheng Y, Lee JD, Kontar RA, Lester C, Yang XJ
The Effects of Presenting AI Uncertainty Information on Pharmacists' Trust in Automated Pill Recognition Technology: Exploratory Mixed Subjects Study
JMIR Hum Factors 2025;12:e60273
URL: https://humanfactors.jmir.org/2025/1/e60273
doi:10.2196/60273
PMID:39932773

Original Paper

Longitudinal and Combined Smartwatch and Ecological Momentary Assessment in Racially Diverse Older Adults: Feasibility, Adherence, and Acceptability Study

Sophia Holmqvist[1], MA; Marina Kaplan[1], BA; Riya Chaturvedi[1]; Haochang Shou[2], PhD; Tania Giovannetti[1], PhD

[1]Department of Psychology and Neuroscience, Temple University, Philadelphia, PA, United States

[2]Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania School of Medicine, Philadelphia, PA, United States

Corresponding Author:
Sophia Holmqvist, MA
Department of Psychology and Neuroscience
Temple University
1701 N. 13th St
Philadelphia, PA, 19122
United States
Phone: 1 215 204 7000
Email: sophia.holmqvist@temple.edu

Abstract

Background: Due to the rising prevalence of Alzheimer disease and related dementias, easily deployable tools to quantify risk are needed. Smartphones and smartwatches enable unobtrusive and continuous monitoring, but there is limited information regarding the feasibility, adherence, and acceptability of digital data collection among racially diverse older adults.

Objective: This paper examined the feasibility, adherence, and acceptability of a 4-week combined smartwatch monitoring and ecological momentary assessment (EMA) study in a racially diverse sample of older adults.

Methods: A total of 44 older adults (aged ≥55 y) with either mild cognitive impairment or healthy cognition completed an informed consent comprehension quiz, baseline cognitive testing, training regarding digital data collection, and questionnaires. Participants were instructed to wear a Garmin Vivosmart 4 smartwatch for 23 h/d for 4 weeks, sync 2 smartphone apps (Garmin and Labfront) daily, and complete a daily EMA survey with automated prompts for surveys and charging. Training time, smartwatch adherence (eg, wear time), daily EMA survey response rate, and performance on the consent quiz were quantified. Associations between feasibility and adherence metrics and participant factors were evaluated. Self-reported usability of the apps and smartwatch was collected at study end.

Results: Consent comprehension quiz scores were high (mean 97.33%, SD 6.86% correct), and training sessions lasted on average 17.93 (SD 6.89) minutes. During the 4-week study, participants wore the smartwatch for an average of 21 h/d (SD 1.53) and showed an average response rate of 94% (SD 9.58%) to daily EMA surveys. In unadjusted bivariate analyses, age, race, and cognition were associated with feasibility and adherence measures, but only age and race remained significant in multivariate models. After accounting for all participant factors, older age was a significant predictor of longer training time, and Black race was a significant predictor of lower daily wear time. On the usability survey, all participants (45/45, 100%) indicated willingness to participate in future smartwatch studies, >80% (37/45) had a positive experience, and >90% (41/45) were satisfied with smartphone app syncing.

Conclusions: Smartwatch monitoring, requiring daily wear, smartphone syncing, and daily EMA survey completion, is highly feasible in older adults because adherence to daily wear and EMA surveys was high, as was general satisfaction on usability surveys. Although older participants may require more training on smartwatch and smartphone procedures and automated prompting during the study period, longitudinal monitoring with the Garmin Vivosmart 4 smartwatch and Labfront app is acceptable and feasible for collecting nearly continuous data in Black and White older adults, including those with mild cognitive impairment and those without.

(JMIR Hum Factors 2025;12:e69952) doi:10.2196/69952

KEYWORDS

cognitive impairment; smartwatch; longitudinal monitoring; ecological momentary assessment; aging

Introduction

Background

With the prevalence of Alzheimer disease (AD) and AD and related dementias (ADRD) increasing alongside the aging population and the availability of new treatments, there is a need to quantify risk and detect cognitive decline at the earliest stage [1]. It is estimated that only a small minority of older adults with mild cognitive decline are accurately identified early [2], missing an opportunity for early intervention. Missed or delayed diagnoses are even more common in Black and African American older adults, who are also at increased risk for AD/ADRD [3]. Personal digital devices such as smartwatches have been proposed as a potential tool for the early detection of AD/ADRD. Although recent data show that smartwatches have been increasingly adopted by older adults [4], the utility of personal digital tools for AD/ADRD monitoring is dependent upon consistent smartwatch use and syncing with associated software. This study examined the feasibility of and adherence to daily smartwatch wear and ecological momentary assessment (EMA) survey completion in a racially diverse sample of older adults enrolled in a 4-week study. We also investigated older adults' comprehension of study details that are relevant to informed consent. The participant characteristics associated with smartwatch adherence and the comprehension of consent were explored to optimize adherence and ethical research conduct in future studies.

Observable yet subtle cognitive, sensory, and motor changes precede and predict the clinical manifestations of AD/ADRD [1,5] and may be measured by sensors in commercially available smartwatches; for example, the Garmin smartwatch has been validated against conventional clinical measures of sleep [6], activity [7,8], and heart rate variability [7], which are associated with dementia risk [9-11]. The Garmin Vivosmart 4 smartwatch also has shown positive results in a 2-week feasibility study of stress in adults undergoing psychotherapy [12]. Thus, digital measures of stress, low heart rate variability, reduced physical activity, and sleep alterations have great potential as low-cost digital biomarkers for AD/ADRD that may be measured continuously, longitudinally, and passively, requiring relatively little effort.

EMA questionnaires delivered directly to participants via their personal digital devices enable high-frequency data collection in the participants' natural setting. EMA can be useful for contextualizing passive metrics [13]; for example, EMA data may facilitate the interpretation of atypical data as clinically significant or not (eg, elevated heart rate due to a cardiac event vs a day at an amusement park). EMA data also provide insight into participants' immediate perceptions and mood states without the confounds of recall bias [14]. There is potential for passive and EMA data collection via personal digital devices, but limited data exist on the feasibility and adherence of device use for health monitoring in older adults. In addition, the extent to which cognitive ability level influences adherence is not known. Adherence to daily EMA surveys and key tasks, such as wearing, charging, and syncing digital devices, is crucial for valid and reliable measurement and requires cognitive resources. This raises an important question: are those who could benefit the most from digital health monitoring more likely to show poor adherence?

Our review of the literature identified few studies on the feasibility and adherence of smartwatch use in digital health research. The results of the available studies suggest several important moderating factors, including the study procedures and participant factors. In-person studies and study designs that include interaction with a member of the study team show much higher retention and adherence than fully remote studies [15]. Adherence also varies depending on task demands, with different adherence rates for smartwatch wear versus syncing versus EMA completion. Adherence to wearable device use, which is relatively less demanding, versus the completion of daily EMA surveys has been estimated to be approximately >90% and 70%, respectively, in studies that are not fully remote [16-19].

Participant factors also play a role in retention and adherence. Although age is commonly considered a barrier to the adoption and acceptability of smartwatches due to low digital literacy [20], empirical data aggregated across several studies of smartwatch data show that older age was associated with greater retention [15]. A study requiring older adults with healthy cognition to wear a Fitbit device for 30 days during waking hours reported high rates of adherence to daily wear (89% of study days) and syncing (85% of study days) [21]. Other studies have demonstrated the effects of race, sex, and memory ability on adherence such that White participants [15], women [22], and people with better memory abilities [21] are more adherent when asked to engage in digital health research studies that involve interacting with a smartphone app.

In addition to the importance of learning about smartwatch adherence and feasibility, it is also crucial to understand older adults' comprehension of informed consent procedures for studies using novel technology. With the large amount of health data collected from a smartwatch and the need for long-term data storage, it is vital for older adult participants to fully understand data privacy and security limitations and the associated risks of study participation. Older adults also may be more wary of sharing personal health information in the context of digital health research. To assess older participants' understanding of informed consent procedures, interactive quizzes have been developed for use in fully remote studies [23]. For this study, which included face-to-face interactions with members of the study team, a 10-item comprehension quiz was modeled after a quiz used by Hackett et al [13] that was designed according to published guidelines [24,25]. The consent quiz focused on the novel technology used in the study (eg, where participants' data are stored and the right to request digital data to be deleted at any time). Hackett et al [13] reported high accuracy rates on a similar quiz in a sample of older adults and significant associations between quiz scores and education level.

Quiz scores also varied by race, with Black participants scoring lower than White participants.

Objectives

In this study, feasibility and adherence were evaluated over a 4-week monitoring period that required wearing the smartwatch daily (23 h/d), charging it, syncing apps, and completing EMA surveys. Feasibility and adherence were assessed by tracking participant retention and measuring the following: (1) time required to complete study training on syncing apps and completing EMA surveys, (2) the comprehension of informed consent information, (3) daily smartwatch wear time, (4) adherence to smartwatch daily wear, and (5) the completion of daily EMA surveys. We also investigated associations between feasibility and adherence measures and participant characteristics (demographics, cognition, and self-reported functional and cognitive decline). Finally, we explored participants' perceptions regarding the usability of the Garmin Vivosmart 4 smartwatch and the study-related apps (syncing process). To address these aims, we recruited a racially diverse sample of older adults, including those with mild cognitive impairment (MCI) and those without. It was hypothesized that older age, worse cognitive function, and greater self-reported functional decline would be associated with lower daily wear time, adherence to wear time, and survey completion, as well as lower comprehension of consent.

Methods

Participants

A total of 47 community-dwelling older adults aged ≥55 years classified as having healthy cognition or MCI were recruited from the Temple University Cognitive Neuropsychology Laboratory cohort in Philadelphia, Pennsylvania. This cohort includes >200 adults who have participated in ongoing federally funded studies since 2020 and represent the racial, ethnic, and economic diversity of the Philadelphia area. Participants met the following inclusion criteria: (1) oral and written fluency in English (to complete study questionnaires and measures) and (2) no history of large vessel stroke, Parkinson disease, major traumatic brain injury, seizures, schizophrenia, or significant neurological conditions other than dementia. The exclusion criteria included (1) current psychiatric disorder (eg, bipolar disorder or major depressive disorder), (2) intellectual disability, and (3) severe motor and sensory deficits precluding the use of a computer touchscreen. Participants from the laboratory cohort were contacted and screened for the following inclusion criteria for this study: (1) aged ≥55 years, (2) cognitive status classified as healthy or MCI, (3) current smartphone user (Android or IOS), and (4) not currently wearing a smartwatch or willing to wear only the study smartwatch during the study period. The exclusion criteria included (1) a diagnosis of dementia and (2) scheduled surgery or travel during the 4-week study period. Approximately 80 participants from the laboratory cohort were contacted as part of the recruitment efforts. The most common reasons for nonparticipation were not having a smartphone or a lack of interest in completing a month-long study.

Of the 47 participants recruited for this study, 1 (2%) was excluded because they required a walker for mobility, which precluded accurate step count recording, and 1 (2%) was excluded from the analyses of feasibility and adherence because their smartphone malfunctioned, and they required the monitoring period to be extended. Moreover, 1 (2%) of the 47 participants dropped out after 1 day, citing discomfort from the smartwatch band being too tight. Thus, of the 47 enrolled participants, 3 (6%) were excluded, resulting in a final sample of 44 (94%) participants for the feasibility and adherence analyses. The participant who extended their monitoring period because of the smartphone malfunction was retained for the usability analyses.

Ethical Considerations

This study was approved by the Temple University Institutional Review Board (Protocol Number: 29712). All participants provided written informed consent and received a US $50 gift card as compensation for their time. Data storage and procedures were Health Insurance Portability and Accountability Act–compliant and only approved study personnel had access to study data.

Procedures

Overview

Participation in the study involved 3 phases. First, participants completed an initial in-person study visit lasting 2 to 4 hours, which included comprehensive cognitive testing, training for the study, and questionnaire completion. Second, participants wore the smartwatch for 4 weeks, answered daily questions on their smartphone, and synced the smartwatch with a smartphone app once a day. Finally, all participants completed a debriefing session. More details of each study phase are provided in the following subsections. Textbox 1 shows the study timeline. The cognitive tests administered during study visit 1 are listed in Table 1.

Textbox 1. Study timeline.

Study visit 1

- Informed consent and comprehension quiz
- Installation of Garmin and Labfront apps
- Study training on syncing and daily survey
- Questionnaires and cognitive testing

4-wk monitoring period

- Participants engage in daily activities as usual
- Wear smartwatch for 23 h/d
- 9-question ecological momentary assessment survey once a day
- Sync Labfront and Garmin apps once a day to securely collect and transmit deidentified data

Study visit 2

- Uninstall apps and unpair smartwatch
- Debriefing questionnaire
- Payment of US $50

Table 1. Cognitive domains and corresponding neuropsychological measures.

Cognitive domains	Neuropsychological measures
Attention	• Trail Making Test part A [26] • Digit Span Forward [27]
Executive function	• Trail Making Test part B [26] • Digit Span Backward [27]
Language	• Letter fluency (S and P) and category fluency (animals) [28] • Boston Naming Test, 30-item version [29]
Processing speed	• Salthouse Letter Comparison Test [30] • Salthouse Pattern Comparison Test [30]
Episodic memory	• Hopkins Verbal Learning Test: immediate recall, delayed recall, and recognition trial [31] • Brief Visual Memory Test–Revised: immediate recall, delayed recall, and recognition trial [32]

Initial Study Visit

Participants completed the following during study visit 1, which lasted 2 to 4 hours: (1) informed consent and 10-item consent comprehension quiz on the specific features of the study (eg, data security and privacy; refer to Multimedia Appendix 1 for the quiz items); (2) gold standard cognitive tests (Table 1) and questionnaires; (3) installation and configuration of study apps and training on wearing a Garmin Vivosmart 4 smartwatch daily (23 h/d of wear; 1 h/d of charging time) for 4 weeks, syncing the smartwatch, and completion of a daily EMA survey. At the end of the first study session, a follow-up visit was scheduled for a date at least 4 weeks later. Further details of the procedures during the initial session are provided in the following subsections.

Smartwatch Data Collection

At the initial study visit, 2 apps (Garmin and Labfront) were downloaded onto the participant's smartphone to facilitate deidentified passive (raw sensor) and active data collection for the daily survey. The Garmin app facilitated data collection from the smartwatch sensors. The Labfront [33] app is a research platform that collects and organizes the Garmin smartwatch data and enabled the study team to monitor adherence and data collection in real time and remotely through a user-friendly study dashboard. The Labfront app was also used to deliver the daily EMA surveys.

A Garmin account was created for each participant using the Garmin Connect mobile app on their smartphone with deidentified information. The Labfront Companion app was then downloaded to the participant's smartphone. The Labfront app uses a randomized 6-digit ID to connect the participant's smartwatch to the app and to connect the Garmin and Labfront accounts. The Garmin smartwatch connects to the Garmin and Labfront apps via Bluetooth. Daily in-app Garmin and Labfront syncing required a Wi-Fi network or an LTE cellular data connection. Once the Garmin and Labfront apps were connected,

data were available for real-time download by the study team in CSV format. Daily syncing of the apps enabled the aggregation of minute-by-minute data.

Participants were asked to wear the Garmin Vivosmart 4 smartwatch for 23 h/d on their nondominant wrist for 4 weeks and to charge it for approximately 1 h/d. Alarms were set in participant smartphones twice daily at times of the participant's choosing to serve as reminders for the following: (1) charging the watch in the morning and (2) syncing the Garmin and Labfront apps and completing the daily survey on the Labfront app at night before going to sleep. To minimize the influence of external cues as well as behavior and health tracking information on participant behavior, all notifications were disabled on the study smartwatch, and additional features were removed from the Garmin app at the initial study visit.

Standardized training procedures with training criteria were used to teach participants how to charge the watch, sync the apps, and complete the daily survey. The training included demonstrations and a practice run during which participants were required to independently complete all daily study steps. Repetitions were recorded for practice runs, and the training session was timed. At the end of the training, participants were given a binder to take home and refer to during the study. The binder contained detailed instructions on how to charge the watch, sync the apps, and complete the daily survey during the study. The contents of this training binder are available in Multimedia Appendix 1.

4-Week Monitoring Period

Participants were told to go about their daily life during the 4-week monitoring period while wearing the smartwatch for 23 h/d and to charge it for 1 h/d. Sleep duration, beat-to-beat heart rate variability, and physical activity data such as step count were collected but are not reported in this paper. Participants were also asked to complete the daily EMA survey delivered to their smartphones through the Labfront app. The EMA survey included 9 questions (Multimedia Appendix 1). The purpose of the EMA survey was to guide the subsequent interpretation of smartwatch data (estimated time to complete the survey: 5 min).

During the course of the study, adherence was monitored through Labfront software by study personnel. If participants were nonadherent for >4 consecutive days, which was defined as wearing the watch for <16.67 h/d or not completing the daily surveys, they were contacted by the study team to determine the reason for nonadherence and to possibly reschedule the follow-up study session to extend the study period and obtain at least 28 days of data collection. The duration of the monitoring period (in days) was tallied for each participant. It is important to note that only 1 (2%) of the 47 enrolled participants required an extension of the monitoring period due to >4 consecutive days of missing data caused by a malfunction of their smartphone. Consequently, this participant was excluded from all feasibility and adherence analyses. For some of the participants (18/44, 41%), the monitoring extended beyond 4 weeks because the follow-up visit could not be scheduled on the study end date due to scheduling conflicts; in these cases, the follow-up visit was scheduled as soon as possible but no more than 1 week later.

Follow-Up Visit

After the study period, participants returned for a brief second visit (study visit 2) during which we collected data on their study experience, the usability of the apps and smartwatch, and barriers to future participation in wearable device research. At this visit, smartwatches were synced a final time to capture any aggregate-level data that had not been previously uploaded to the app. Subsequently, deidentified data obtained during the monitoring period were downloaded in CSV format for each participant from the Labfront researcher user interface for processing and analysis.

Measures of Participant Characteristics

Cognitive tests and questionnaires used in standard clinical evaluations for cognitive decline were administered during study visit 1.

Cognition and Clinical Classification

IQ was estimated with the Hopkins Reading Test [34]. The 5 cognitive domains and 10 cognitive tests that were administered are listed in Table 1. Scores from each of the 10 tests were standardized (T score) after adjusting for demographic variables (age, education, sex, and estimated IQ score [28]). Clinical classification (ie, healthy cognition vs MCI) was based on published criteria [35] that define MCI as T scores of ≤40 (<1 SD) on both measures from at least 1 cognitive domain. A total composite score and cognitive domain–specific composite scores were calculated by averaging demographically corrected T scores (refer to Table 1 for all neuropsychological measures). These neuropsychological tests are used widely in research and clinic settings, have been extensively validated, and show strong psychometric properties [28].

Self-Reported Cognitive and Functional Decline

The average score from the Everyday Cognition (ECog)–short form [36] was used to estimate self-reported functional decline across a range of abilities (eg, memory and language), with scores ranging from 1 (no decline over the past 10 y) to 4 (worse decline over the past 10 y). The optimal cut score for distinguishing between healthy cognition and MCI is 1.32. The original validation study estimated a Cronbach α value of 0.96 [36].

Smartwatch Feasibility and Adherence Measures

Training Time

The amount of time spent training (in minutes) during study visit 1 was recorded. Training procedures were standardized (Multimedia Appendix 1). Training began with a review of procedures for charging the watch and ended when the examiner observed that the participant independently demonstrated charging the watch, syncing the apps, and completing the daily survey.

Comprehension of Consent

Comprehension of consent was assessed using a 10-item quiz (Multimedia Appendix 1), which included questions on study-specific risks, data privacy, and security. The following is an example item: "The Garmin/Labfront apps collect the *content* of my texts and phone calls" (correct answer: no). To

ensure informed consent, if participants answered any item incorrectly, study personnel immediately provided feedback to clarify the correct response.

Smartwatch Adherence

The total number of days that participants did not wear the watch was tallied, and 2 measures were used to evaluate adherence to the instruction to wear the watch for 23 h/d. First, the average daily wear time (in hours) per participant across all possible days of wear was collected. Wear time was derived from minute-to-minute heart rate data; therefore, wear time reflects the accumulation of heart rate data from the watch. Weekly averages of wear time were also computed (weeks 1, 2, 3, and 4) to examine adherence over time.

On the basis of published recommendations [37], daily watch data were considered valid if the participant wore the smartwatch for at least 16.67 hours that day. Therefore, watch adherence also was assessed as the "percentage of valid smartwatch days" during the monitoring period, calculated as the number of days the participants wore the watch for ≥16.67 h/d divided by the total number of days they were in the study.

EMA Survey Adherence

Adherence to the daily survey was calculated as the number of daily surveys completed divided by the total number of days that each participant was in the study. This percentage ranged from 0% to 100%, with 100% indicating perfect adherence.

Usability Measures

A self-administered usability survey was completed by participants at the follow-up visit (study visit 2) to provide qualitative information on smartwatch use. The survey questions focused on the smartwatch and Labfront and Garmin smartphone apps, which were used for daily EMA surveys and syncing. Participants were also asked to report their likelihood of participating in another future study that included the smartwatch and daily surveys. This survey was administered to 45 (96%) of the 47 enrolled participants, including the participant whose monitoring period was extended.

Statistical Analyses

Descriptive statistics were used to characterize feasibility: average daily wear time, study training time, percentage correct on consent comprehension, the percentage of completed EMA surveys, the percentage of valid smartwatch days, and responses to the study debriefing questionnaire. To examine univariate associations between feasibility and adherence measures and participant characteristics, Spearman correlations and Mann-Whitney *U* tests were used for continuous and categorical variables, respectively, with Bonferroni correction applied to interpret statistical significance. Friedman rank tests for 1-way repeated measures of ANOVA were used to analyze whether wear time differed from week 1 to week 4, with Bonferroni correction applied to interpret statistical significance. Nonparametric statistical tests were used due to the skewed distributions of the variables of interest. Multiple linear regressions were conducted to examine multivariate associations between demographics (ie, age, sex, race, and education) and cognitive variables (ie, ECog, cognitive composite scores, and cognitive status) on all feasibility and adherence measures. Separate models were conducted for each cognitive variable to avoid multicollinearity. As the dependent variables were not normally distributed, we used bootstrapping (5000 replications) to estimate robust SEs and CIs for coefficients, leading to more reliable testing of significance.

All analyses were conducted using SPSS software (version 28.0; IBM Corp) or, for bootstrapped regressions only, R (version 2023.12.1+402; R Foundation for Statistical Computing).

Results

Participant Characteristics

A total of 44 older adults ranging in age from 55 to 83 years participated and were included in the feasibility and adherence analyses. On average, participants were college educated with estimated IQ scores in the high average range (Table 2). Self-report of cognitive and functional decline was well within the normal range (ECog; range 1-2.5). The cognitive composite T score ranged from 35.40 to 63.40. The sample comprised mostly women (26/44, 59%) and primarily individuals identifying as Black or African American (11/44, 25%) or White, non-Hispanic (31/44, 70%). Of the 44 participants, 8 (18%) met the Jak-Bondi criteria for MCI. The participants were relatively active, with an average of 5999.74 (SD 2872.38) steps per day. Of the 44 participants, 17 (39%) owned a smartwatch before the study, with most owning a Fitbit device (n=9, 50%) or an Apple Watch (n=9, 50%). Those who currently or previously owned a smartwatch reported the main motivations for use as activity monitoring, followed by improving fitness and improving health. Of those who had not owned a smartwatch, most (17/44, 39%) indicated that they would consider buying one after their study experience.

Table 2. Participant characteristics and smartwatch feasibility measures (n=44).

Variables	Values
Participant characteristics	
Age (y), mean (SD)	68.48 (7.22)
Education (y), mean (SD)	16.41 (2.08)
Sex: female, n (%)	26 (59)
Race and ethnicity, n (%)	
Asian	1 (2)
Black or African American	11 (25)
White, Hispanic	1 (2)
White, non-Hispanic	31 (70)
MCI[a], n (%)	8 (18)
Estimated IQ, mean (SD)	110.84 (7.59)
ECog[b] score, mean (SD)	1.41 (0.39)
Cognitive composite score, mean (SD)	51.69 (5.68)
Daily step count, mean (SD)	5999.74 (2872.38)
Smartwatch ownership before study, n (%)	17 (39)
"If you already currently own and use a smartwatch or have previously, what is or was your main motivation for using the smartwatch? Select all that apply," n (%)	
To monitor activities	15 (34)
Improve fitness	15 (34)
Improve health	12 (27)
Keep up with new technology	7 (16)
Improve appearance	2 (4)
"If you do not already own a smartwatch, would you consider buying one after this experience?" n (%)	
Yes	17 (39)
No	10 (24)
Already have smartwatch	17 (39)
Smartwatch feasibility and adherence measures, mean (SD)	
Percentage correct of consent comprehension	98.00 (6.86)
Average training time (min)	17.93 (6.89)
Average daily smartwatch wear time (h)	21.04 (1.53)
Percentage of valid smartwatch days	92.00 (10.39)
Percentage of completed daily surveys	94.00 (9.58)

[a]MCI: mild cognitive impairment.

[b]ECog: Everyday Cognition.

Retention Rate and Duration of Monitoring Period

A total of 47 participants were enrolled, of whom 1 (2%) dropped out after 1 day citing discomfort from the smartwatch band being too tight. All other participants (46/47, 98%) completed the study. However, 2 (4%) of these 46 participants were excluded from feasibility data analysis: 1 (50%) was excluded because their monitoring period was extended to 46 days due to a smartphone malfunction, and they did not complete the EMA survey for 1.5 weeks during the originally scheduled 28-day study; consequently, they were asked and agreed to extend their monitoring period. As a result, this participant was excluded from the analytic sample used for feasibility and adherence analyses. No other participants extended their monitoring period for reasons other than scheduling conflicts. Another participant was excluded due to mobility issues. Thus, the final analytic sample for the feasibility analyses comprised 44 participants.

The number of days monitored with the smartwatch ranged from 27 to 35. The mode for study days was 27 days,

corresponding to exactly 4 weeks because the date of the initial study visit was excluded from analyses. The average number of days in the study was 29 (SD 3.46) days. Feasibility and adherence measures were computed using all available monitoring days.

Feasibility and Adherence

Average feasibility and adherence measures are shown at the bottom of Table 2.

Comprehension Quiz

On average, participants obtained a score of 97% (range 70%-100%) correct on the consent comprehension quiz, approaching a perfect average score. The most frequently missed response was indicating "yes" to "Using the Garmin and Labfront apps will help improve my cognitive functioning."

Training Time

Participants took an average of 17.93 (range 12-53) minutes to complete the study training during study visit 1. Training time was mostly spent independently practicing study tasks, such as syncing apps.

Smartwatch Adherence

Of the 44 participants, 11 (25%) did not wear the watch for at least 1 entire day (n=9, 20% missed only 1 d; n=1, 2% missed 2 d; and n=1, 2% missed 4 d). All participants wore the watch for ≥16.67 h/d for ≥14 days during the monitoring period. Of the 44 participants, 5 (11%) had ≥27 days of wear time that exceeded 16.67 h/d. On average, participants exceeded 16.67 h/d of wear time on 92% (range 53%-100%) of total study days. As the monitoring period extended beyond 4 weeks for some participants due to scheduling conflicts, watch adherence measures also were calculated over 27 days, with the average percentage of days exceeding 16.67 h/d of wear time remaining unchanged at 92% (range 55%-100%).

On average, participants wore the watch for 21 (range 15.71-22.44) h/d across the 4-week study period. Of the 44 participants, 3 (7%) had an average wear time of <18 h/d, 2 (5%) had an average wear time of <16.67 h/d, and 39 (89%) wore the watch for ≥20 h/d on average.

The distribution of average watch wear time per week is shown in Figure 1. Pairwise 2-tailed *t* tests showed no significant difference in average wear time across the 4 weeks (refer to Multimedia Appendix 1 for full results).

As shown in Figure 2, on average, 91% (40/44) of the participants wore the watch for at least 19 h/d, within 4 hours of the required 23 h/d. In fact, on average, participants on average wore the watch for 21 h/d, within 2 hours of the required 23 h/d.

Figure 1. Daily wear time across participants was averaged across weeks 1 to 4 in the study. The y-axis represents the mean daily smartwatch wear time in hours. The error bars represent –1 to +1 SD.

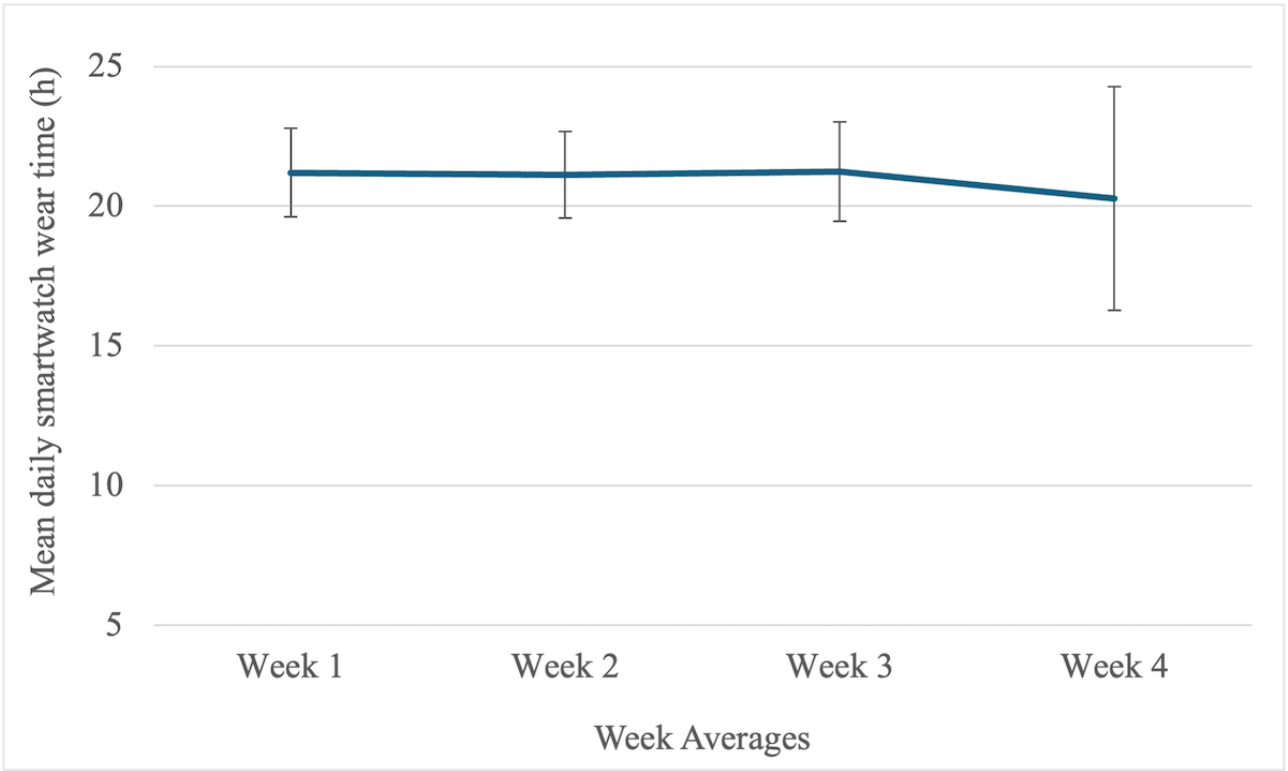
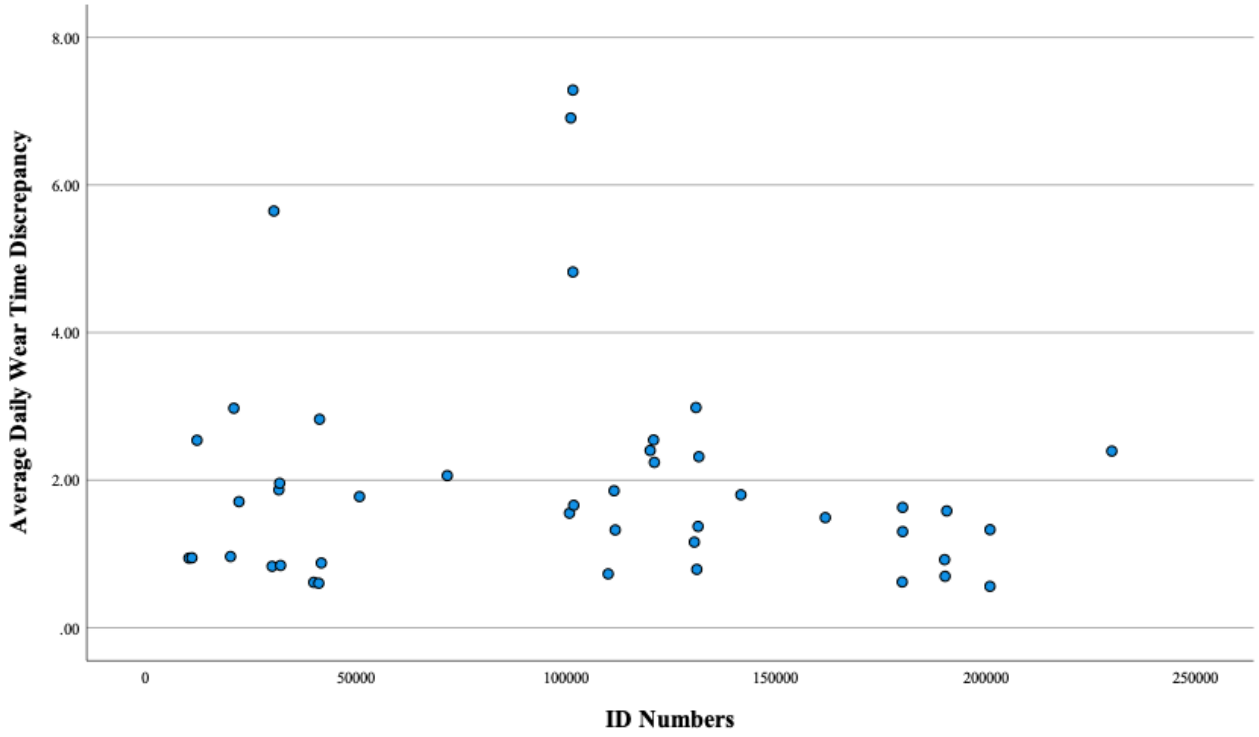


Figure 2. Average wear time discrepancy for the entire sample. The y-axis represents each participant's average daily wear time subtracted from the required 23 h/d. The x-axis shows participant ID numbers from the feasibility analyses. Each dot represents a participant.



EMA Survey Adherence

On average, participants completed surveys on most study days (94%; range 58%-100%). Of the 44 participants, 3 (7%) completed <70% of the surveys, while 41 (93%) completed >85% of the daily surveys.

Correlations Among Smartwatch Feasibility and Adherence Metrics

As shown in Table 3, a greater percentage of daily surveys completed was significantly associated with longer study training time, but the *P* value for this association did not survive Bonferroni correction (ie, *P*=.005 and below). Correlations between the 2 measures of smartwatch adherence and between smartwatch adherence measures and the percentage of daily surveys completed were significant and survived correction for multiple comparisons. The correlations indicated that participants who completed a higher percentage of daily surveys also wore the watch for longer durations and had more days of valid smartwatch wear time.

Table 3. Correlation coefficients among smartwatch feasibility and adherence metrics.

Variables	Consent comprehension percentage correct	Study training time	Percentage of daily surveys completed	Percentage of valid smartwatch wear days
Study training time				
r	0.055	—[a]	—	—
P value	.77	—	—	—
Percentage of daily surveys completed				
r	–0.016	0.298	—	—
P value	.92	.049	—	—
Percentage of valid smartwatch wear days				
r	0.230	0.018	0.544	—
P value	.13	.91	<.001	—
Daily wear time				
r	0.095	0.091	0.579	0.778
P value	.54	.56	<.001	<.001

[a]Not applicable.

Associations Between Participant Characteristics and Feasibility and Adherence Metrics

As shown in Table 4, only age and self-reported cognitive and functional decline (ECog) were significantly associated with feasibility and adherence metrics. Specifically, older participants required longer study training times, and participants who reported greater cognitive and functional decline (ECog) had lower daily average wear time. However, these significant correlations did not survive Bonferroni correction (P=.003). No other correlation coefficients were statistically significant (Table 4).

Table 4. Correlation coefficients between smartwatch feasibility metrics and demographic and clinical factors.

Variables	Consent comprehension percentage correct	Study training time	Percentage of daily surveys completed	Percentage of valid smartwatch wear days	Daily wear time
Age					
r	–0.164	0.403	0.165	–0.023	0.006
P value	.29	.007	.28	.88	.97
Education					
r	0.041	0.029	–0.197	–0.092	–0.108
P value	.79	.85	.20	.55	.48
Cognitive composite scores					
r	0.216	–0.002	0.037	0.102	0.155
P value	.16	.99	.81	.51	.31
ECog[a]					
r	–0.102	–0.206	–0.221	–0.261	–0.393
P value	.51	.19	.15	.09	.009

[a]ECog: Everyday Cognition.

There were no significant sex differences in consent comprehension, training time, watch adherence measures, or the percentage of daily surveys completed (data reported in Multimedia Appendix 1). Participants with MCI and those with healthy cognition differed on only 1 measure (data reported in Multimedia Appendix 1): participants with MCI (median 20.64, IQR 2.74 h) had significantly lower daily wear time than participants with healthy cognition (median 21.53, IQR 1.18 h; U=71.00; P=.03). This result is consistent with the significant correlation between ECog and daily wear time reported in Table 4, suggesting an association between lower cognitive abilities and less smartwatch wear time.

Between-group analyses comparing Black and White participants were not statistically significant, except for daily wear time (data reported in Multimedia Appendix 1). Black participants (median 20.76, IQR 0.91 h) had significantly lower daily smartwatch wear time than White participants (median 21.83, IQR 1.02 h; U=74.00; P=.004).

Multivariate associations between participant characteristics and feasibility and adherence metrics were investigated using multiple regression analyses with bootstrapping to estimate the significance of the coefficients. Separate regressions were run, including one for each cognitive variable (cognitive composite scores, ECog, and cognitive status; refer to Multimedia Appendix 1 for all tables), to avoid multicollinearity. The results showed that race remained a significant predictor of daily wear time, even after controlling for age, sex, education, and measures of cognition, with White participants having longer daily wear times than Black participants. Multivariate analyses also showed that age was significantly associated with training time even after controlling for sex, education, race, and measures of cognition, with older age associated with longer study training time. By contrast, associations between the measures of cognition and feasibility and adherence metrics were not statistically significant in regression analyses, including demographic variables (age, sex, education, and race).

Usability Survey

The results of the usability survey are shown in Table 5 (Garmin smartwatch) and Table 6 (Labfront smartphone app). Overall, participants expressed favorable views of both the smartwatch and the smartphone apps and especially enjoyed their involvement in research. When asked, “Would you wear this Garmin smartwatch without being asked to wear it as part of the study?” 55% (25/45) of the sample reported “yes.” However, all participants indicated they were likely (11/45, 24%) or very likely (34/45, 76%) to participate in a future study asking them to wear a Garmin smartwatch daily. Of the 45 participants, 36 (80%) either somewhat or strongly agreed that they had a positive experience using the Garmin smartwatch during the study period (Table 5). Most of the participants (31/45, 68%) reported that the smartwatch was comfortable or very comfortable. Most of the participants (32/45, 71%) were satisfied or very satisfied with the functioning of the watch. The majority of the complaints about the smartwatch were regarding low battery life (17/45, 38%), followed by discomfort (8/45, 18%) and technical issues (8/45, 18%). Of the 45 participants, 17 (38%) reported no complaints about the smartwatch. Participants suggested the following improvements for the watch: changing wristband material (18/45, 40%), increasing

the display size (15/45, 33%), and increasing the wristband size (10/45, 22%).

Table 5. Smartwatch usability and satisfaction (n=45).

Survey items	Participants, n (%)
"How comfortable was the smartwatch?"	
Very comfortable[a]	20 (44)
Comfortable	11 (24)
Somewhat comfortable	10 (22)
Not comfortable	4 (9)
"Did you wear it overnight every night?"	
Yes	44 (98)
"How satisfied were you with the functioning of the watch (i.e. you were able to tell date/time easily)?"	
Very satisfied	18 (40)
Satisfied	14 (31)
Somewhat satisfied	9 (20)
Not satisfied	4 (9)
"Do you have any complaints about using the Garmin Vivosmart 4 watch?"	
None	17 (38)
Technical issues	8 (18)
Doesn't fit	2 (4)
It's uncomfortable	8 (18)
Low battery life	17 (38)
Problems with the screen	6 (13)
"Overall, I had a positive experience using the smartwatch."	
Disagree strongly	0 (0)
Somewhat disagree	3 (7)
Neutral	5 (11)
Somewhat agree	18 (39)
Agree strongly	19 (43)
"What would you change to improve the comfort of wearing the Garmin smartwatch?"	
No changes needed	15 (33)
Improve wristband clasp function	9 (20)
Increase display size	15 (33)
Change material of wristband (cloth, material)	18 (40)
Other (increase size of wristband)	10 (22)
"How satisfied were you with charging the battery of the Garmin smartwatch?"	
Very satisfied	10 (22)
Satisfied	19 (44)
Somewhat satisfied	11 (24)
Not satisfied	4 (9)
"Did the watch ever run out of battery while you were wearing it?"	
Yes (those who answered yes, indicated between 1 and 5 times)	17 (38)

[a]Italicized text indicates the most frequent response.

App usability and overall experience are reported in Table 6. Of the 45 participants, 41 (91%) were satisfied or very satisfied with the syncing process of both apps (Garmin and Labfront). Of the 45 participants, 35 (78%) indicated that they agreed or strongly agreed that the Labfront app was easy to use and that they felt very confident using it. Importantly, 64% (29/45) of the participants had a positive or extremely positive experience with the Labfront app. Moreover, 71% (32/45) of the participants indicated that the Labfront app was easy to use or extremely easy to use (Table 6).

Table 6. App usability and overall experience (n=45).

Survey items	Participants, n (%)
"How easy was it to sync the watch every day with the Labfront and Garmin apps?"	
Very satisfied[a]	32 (71)
Satisfied	9 (20)
Somewhat satisfied	1 (2)
Not satisfied (issues included having to try a couple of times to sync the Labfront app)	3 (7)
"I thought Labfront was easy to use"	
Strongly disagree	3 (7)
Disagree	5 (11)
Neither agree nor disagree	2 (4)
Agree	18 (40)
Strongly agree	17 (37)
"I felt very confident using the Labfront app"	
Strongly disagree	2 (4)
Disagree	2 (4)
Neither agree nor disagree	4 (9)
Agree	18 (40)
Strongly agree	19 (42)
"I enjoy using the Labfront app"	
Strongly disagree	1 (2)
Disagree	2 (4)
Neither agree nor disagree	18 (40)
Agree	17 (38)
Strongly agree	7 (16)
"How would you rate your experience with the Labfront app?"	
Extremely negative	1 (2)
Negative	2 (4)
Neutral	13 (29)
Positive	23 (51)
Extremely positive	6 (13)
"How would you rate usability of the Labfront app?"	
Very difficult to use	0 (0)
Somewhat difficult to use	3 (7)
Neutral	10 (22)
Easy to use	23 (53)
Extremely easy to use	8 (18)

[a]Italicized text indicates the most frequent response.

Discussion

Principal Findings

Overall, the results demonstrated that our longitudinal study requiring daily wear of a smartwatch and daily EMA survey completion in a racially diverse sample of older adults was feasible with excellent adherence. Participants with cognitive impairment and those without wore a commercially available smartwatch during waking and sleep hours for 4 weeks for an average of 21 h/d and showed an average response rate of 94% to daily EMA surveys. Only 1 (2%) of the 47 enrolled participants refused to complete the study after informed consent due to discomfort with the watch fit. On the basis of the results of our consent comprehension quiz, participants had no difficulty understanding the study procedures and risks, and completed smartwatch and app training in 18 minutes on average. After consent and training, participants remained in the study for at least 28 days. Average wear time did not decline significantly over the course of the study and on average never fell below 20 h/d. Although there are no benchmark standards for indicating good smartwatch adherence, a recent systematic review of activity trackers to monitor physical activity suggested that 3 valid days of at least 10 h/d may be a good adherence threshold for a week-long study [38]. Another study defined a "valid day" as one that includes at least 16.67 hours of data or at least 600 one-minute epochs of nonzero heart rate values [37]. In our study, participants wore the study watch for >16.67 h/d on 92% of the days on average. Thus, our results exceed the recent suggestions for adherence and validity and show strong support for the feasibility and adherence of our protocol to monitor older adults.

Smartwatch feasibility and adherence metrics were associated with each other as expected, and EMA survey adherence also was related to smartwatch adherence measures. However, feasibility and adherence metrics were differentially related to participant factors, indicating the importance of measuring adherence to each feature of a smartwatch and EMA study separately. EMA survey adherence was not related to any participant characteristic and was uniformly high, possibly because participants received smartphone reminder alarms to complete the daily survey and sync the apps at predetermined times at night before bedtime. There was no comparable alert to prompt participants to wear the smartwatch during the day or night. Future studies should consider adding prompts to facilitate smartwatch adherence.

In contrast to past studies [21], we observed no associations between participant's sex or education level on any measure of adherence and feasibility in our diverse sample. Smartwatch adherence was associated with race such that White participants wore the smartwatch for more hours per day (daily wear time) than Black participants during the study period. The effect of race remained significant, even after controlling for other demographic variables, including age, education, and cognition. The reason for the race difference is unknown but may be explained by different beliefs, attitudes, or daily habits between the racial groups. As wear time is measured using the photoplethysmography sensor, which relies on the absorption of infrared light into the skin, it is possible that the measures might have been less accurate for people with darker skin tone. If replicated, cultural factors and potential sensor limitations related to skin tone, which might explain racial differences in adherence, should be systematically explored. Although Black and White participants differed in daily wear time, race did not influence the percentage of days that reached the 16.67-hour threshold for "valid" wear time. Thus, the race difference in daily wear time was small and potentially inconsequential. Furthermore, race was not associated with any other feasibility and adherence measure.

Although cognitive ability level (both self-reported cognitive decline and MCI status) was associated with daily wear time in bivariate analyses, the effect of cognition was not significant in multiple regression models. These results contrast with past research that has shown cognitive ability to be related to adherence; for example, memory ability was related to adherence in a study requiring the daily syncing of a Fitbit device in a sample of older adults without cognitive impairment [21]. Our study protocol, including training and supports during the monitoring period (eg, the take-home binder and twice daily alarms [in the morning to charge the watch and at night to sync the apps and complete the surveys]) may have helped participants with cognitive difficulties adhere to the smartwatch wear requirements and EMA survey.

Age was not associated with watch or survey adherence, which is consistent with the results from a recent smartphone digital phenotyping study [15]. However, similar to a recent study from our laboratory [13], we found that age was associated with training time such that older participants required longer training times in the laboratory to learn how to sync the study-related apps and complete the daily EMA survey on their smartphone. The training required participants to independently complete each activity (syncing apps, charging the smartwatch, and completing the EMA survey) while a member of the study team observed, answered questions, and provided feedback. On average, participants took <20 minutes to complete the training. Older participants may have taken longer to complete the training due to lower digital literacy; they may have required more practice and may have asked more questions during the training. Although the number of people aged >65 years owning smartphones in the United States has steadily increased over the past decade, digital literacy is generally lower in older adults. The longer training time and the take-home binder, which summarized the training for review during the study period, may have consequently minimized age effects on smartwatch and daily EMA survey adherence. Smartwatch studies that incorporate a "human in the loop" have been found to have higher adherence than fully remote studies [15]. Even if a study is conducted fully remotely, providing some contact with study personnel (via teleconference) or training resources (in the form of a training binder, video tutorial, website, or web-based helpline) could be quite beneficial to increasing adherence, especially for older adults with low digital literacy.

Participants generally demonstrated solid knowledge of the informed consent information, including issues related to data security and privacy that are relevant to research using wearable devices; for example, responses to the study consent quiz

showed that participants understood that the study apps were Health Insurance Portability and Accountability Act–compliant and that they had control over their data such that they had the right to delete their study data at any time. In contrast to a study on smartphone digital phenotyping that found that education and race were associated with quiz accuracy [13], percentage correct on the consent comprehension quiz was not significantly related to any of the participant demographic factors, possibly because the scores on the quiz had little variability (ie, ceiling effects). Similar to a past study [13], the most frequently incorrect question concerned the potential benefits of the study, with many participants incorrectly reporting that their participation would improve their cognitive abilities. In these cases, the comprehension quiz enabled the study team to explain that wearing the smartwatch had no clear benefit and possibly preclude participants' disappointment at the end of the study.

Smartwatch usability reports were largely very positive, with most of the participants (37/45, 82%) agreeing that they had a positive experience using the Garmin Vivosmart 4 smartwatch in the study as well as high satisfaction regarding its comfort and functioning. This is in line with past research showing that the Garmin Vivosmart was rated as the most usable and acceptable smartwatch compared to 5 other smartwatches by older adults in a small study [39]. Although participants indicated that they needed to charge the Garmin Vivosmart 4 smartwatch daily, they were generally satisfied with battery life. Regarding app usability, almost all participants were satisfied with the syncing process of the Labfront and Garmin Connect apps (42/45, 93%), felt that the Labfront app was easy to use (35/45, 78%), and reported a positive experience using the Labfront app (29/45, 64%). There are currently no other studies examining the usability and acceptability of the Labfront and Garmin Connect apps.

Regarding the usability of the smartwatch and apps, participants were generally satisfied with the smartwatch and reported that they would wear it again in another study. It is important to consider that most of the participants (44/45, 98%) reported having to charge the watch daily, and low battery life was the most frequently cited complaint about the smartwatch. Thus, in future studies, participants should be trained and prompted to charge watches and other digital devices, and they should be informed about battery limitations to preclude complaints and dissatisfaction. Multiple chargers, including portable batteries, may be offered to participants to improve usability. Despite minor complaints regarding charging, most of the participants (37/45, 82%) reported a positive experience during the study using the smartwatch. Participants also reported high satisfaction and ease of use of the Labfront and Garmin apps, which were used for daily syncing to enable sensor and EMA data to be transmitted to the research team. Taken together, our findings are generally in line with prior research showing good acceptability and usability of wearable devices in research studies involving older adults [21].

Limitations

Although our sample included older adults of different races, the sample was relatively small and highly educated. Furthermore, only 8 (18%) of the 44 participants had MCI; therefore, analyses comparing participants with MCI and those with healthy cognition were underpowered. However, cognitive effects on adherence were also examined with a composite test score and self-reports of cognitive and functional abilities (ECog). Moreover, our monitoring period was limited to a 4-week period; therefore, the conclusions may not generalize to longer periods. Importantly, our study incorporated several supports for participants such as in-person training, a take-home binder, contact with study personnel, and reminder alarms set for times specified by participants. These study features likely influenced the high adherence to smartwatch wear and EMA survey completion in our study compared to fully remote studies with little to no support from study personnel. In addition, our study's EMA survey imposed minimal demands because it was required only once daily and completed at a time preferred by participants.

Conclusions

This study demonstrated the feasibility of a combined smartwatch monitoring and daily EMA study requiring nearly 24-hour wear time daily for a month in racially diverse older adults, including those with cognitive impairment and those without. Smartwatch feasibility and adherence metrics were not significantly intercorrelated and were differentially related to participant characteristics. The inclusion of a guided training for older adult participants on study procedures as well as a take-home binder and smartphone reminder alarms to complete study steps (syncing apps, charging the smartwatch, and completing the daily EMA survey) may have contributed to strong adherence and feasibility in this study, regardless of cognitive impairment. Participants reported high satisfaction regarding the usability of both the Garmin Vivosmart 4 smartwatch and Labfront app. Overall, longitudinal monitoring using a commercially available smartwatch with a daily EMA survey is acceptable and feasible for collecting nearly continuous data in Black and White older adults, including those with cognitive impairment and those without.

Acknowledgments

The research in this publication was supported by the National Institute on Aging (F31AG089944: SH; R01AG062503: TG). The publication of this article was funded in part by the Temple University Libraries Open Access Publishing Fund.

Authors' Contributions

SH conceptualized the idea, ran statistical analyses, and drafted the manuscript. RC and MK contributed to data processing, analysis, and editing. TG contributed to conceptualization, writing, and editing. HS contributed to editing and statistical analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Training binder, consent comprehension quiz, and supplementary analyses.
[PDF File (Adobe PDF File), 1432 KB - humanfactors_v12i1e69952_app1.pdf]

References

1. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/ wearable devices opportunity. NPJ Digit Med 2019;2:9. [doi: 10.1038/s41746-019-0084-2] [Medline: 31119198]
2. Frederiksen KS, Arus XM, Zetterberg H, Gauthier S, Boada M, Pytel V, et al. Focusing on earlier diagnosis of Alzheimer's disease. Future Neurol 2024 Mar 27;19(1):33. [doi: 10.2217/fnl-2023-0024]
3. Lin PJ, Daly AT, Olchanski N, Cohen JT, Neumann PJ, Faul JD, et al. Dementia diagnosis disparities by race and ethnicity. Med Care 2021 Aug 01;59(8):679-686 [FREE Full text] [doi: 10.1097/MLR.0000000000001577] [Medline: 34091580]
4. Faverio M. Share of those 65 and older who are tech users has grown in the past decade. Pew Rearch Centre. URL: https://www.pewresearch.org/short-reads/2022/01/13/share-of-those-65-and-older-who-are-tech-users-has-grown-in-the-past-decade/ [accessed 2024-09-11]
5. Sperling R, Mormino E, Johnson K. The evolution of preclinical Alzheimer's disease: implications for prevention trials. Neuron 2014 Dec 05;84(3):608-622 [FREE Full text] [doi: 10.1016/j.neuron.2014.10.038] [Medline: 25442939]
6. Schyvens AM, Van Oost NC, Aerts JM, Masci F, Peters B, Neven A, et al. Accuracy of Fitbit Charge 4, Garmin Vivosmart 4, and WHOOP versus polysomnography: systematic review. JMIR Mhealth Uhealth 2024 Mar 27;12:e52192 [FREE Full text] [doi: 10.2196/52192] [Medline: 38557808]
7. Evenson KR, Spade CL. Review of validity and reliability of Garmin activity trackers. J Meas Phys Behav 2020 Jul;3(2):170-185 [FREE Full text] [doi: 10.1123/jmpb.2019-0035] [Medline: 32601613]
8. Briggs BC, Hall KS, Jain C, Macrea M, Morey MC, Oursler KK. Assessing moderate to vigorous physical activity in older adults: validity of a commercial activity tracker. Front Sports Act Living 2021 Jan 3;3:766317 [FREE Full text] [doi: 10.3389/fspor.2021.766317] [Medline: 35047769]
9. Cheng YC, Huang YC, Huang WL. Heart rate variability in patients with dementia or neurocognitive disorders: a systematic review and meta-analysis. Aust N Z J Psychiatry 2022 Jan 08;56(1):16-27. [doi: 10.1177/0004867420976853] [Medline: 33287558]
10. Fan L, Xu W, Cai Y, Hu Y, Wu C. Sleep duration and the risk of dementia: a systematic review and meta-analysis of prospective cohort studies. J Am Med Dir Assoc 2019 Dec;20(12):1480-7.e5. [doi: 10.1016/j.jamda.2019.06.009] [Medline: 31604673]
11. Tan ZS, Spartano NL, Beiser AS, DeCarli C, Auerbach SH, Vasan RS, et al. Physical activity, brain volume, and dementia risk: the Framingham study. J Gerontol A Biol Sci Med Sci 2017 Jul 01;72(6):789-795 [FREE Full text] [doi: 10.1093/gerona/glw130] [Medline: 27422439]
12. Hehlmann MI, Schwartz B, Lutz T, Gómez Penedo JM, Rubel JA, Lutz W. The use of digitally assessed stress levels to model change processes in CBT - a feasibility study on seven case examples. Front Psychiatry 2021;12:613085 [FREE Full text] [doi: 10.3389/fpsyt.2021.613085] [Medline: 33767638]
13. Hackett K, Xu S, McKniff M, Paglia L, Barnett I, Giovannetti T. Mobility-based smartphone digital phenotypes for unobtrusively capturing everyday cognition, mood, and community life-space in older adults: feasibility, acceptability, and preliminary validity study. JMIR Hum Factors 2024 Dec 22;11:e59974 [FREE Full text] [doi: 10.2196/59974] [Medline: 39576984]
14. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
15. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
16. Goodday SM, Karlin E, Alfarano A, Brooks A, Chapman C, Desille R, et al. An Alternative to the Light Touch Digital Health Remote Study: The Stress and Recovery in Frontline COVID-19 Health Care Workers Study. JMIR Form Res 2021 Dec 10;5(12):e32165 [FREE Full text] [doi: 10.2196/32165] [Medline: 34726607]
17. Druce KL, McBeth J, van der Veer SN, Selby DA, Vidgen B, Georgatzis K, et al. Recruitment and ongoing engagement in a UK smartphone study examining the association between weather and pain: cohort study. JMIR Mhealth Uhealth 2017 Dec 01;5(11):e168 [FREE Full text] [doi: 10.2196/mhealth.8162] [Medline: 29092810]
18. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. JMIR Mhealth Uhealth 2017 Mar 24;5(3):e37 [FREE Full text] [doi: 10.2196/mhealth.6496] [Medline: 28341616]

19. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, RADAR-CNS consortium. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Mar 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
20. Baig MM, Afifi S, GholamHosseini H, Mirza F. A systematic review of wearable sensors and IoT-based monitoring applications for older adults - a focus on ageing population and independent living. J Med Syst 2019 Jul 15;43(8):233. [doi: 10.1007/s10916-019-1365-7] [Medline: 31203472]
21. Paolillo EW, Lee SY, VandeBunte A, Djukic N, Fonseca C, Kramer JH, et al. Wearable use in an observational study among older adults: adherence, feasibility, and effects of clinicodemographic factors. Front Digit Health 2022 Jun 10;4:884208 [FREE Full text] [doi: 10.3389/fdgth.2022.884208] [Medline: 35754462]
22. Pathiravasan CH, Zhang Y, Wang X, Trinquart L, Benjamin EJ, Borrelli B, et al. Factors associated with long-term use of digital devices in the electronic Framingham Heart Study. NPJ Digit Med 2022 Dec 27;5(1):195 [FREE Full text] [doi: 10.1038/s41746-022-00735-1] [Medline: 36572707]
23. McInnis BJ, Pindus R, Kareem D, Gamboa S, Nebeker C. Exploring the future of informed consent: applying a service design approach. Proc ACM Hum Comput Interact 2024 Apr 26;8(CSCW1):1-31. [doi: 10.1145/3637330]
24. Bartlett Ellis R, Wright J, Miller LS, Jake-Schoffman D, Hekler EB, Goldstein CM, et al. Lessons learned: beta-testing the digital health checklist for researchers prompts a call to action by behavioral scientists. J Med Internet Res 2021 Dec 22;23(12):e25414 [FREE Full text] [doi: 10.2196/25414] [Medline: 34941548]
25. Nebeker C, Bartlett Ellis RJ, Torous J. Development of a decision-making checklist tool to support technology selection in digital health research. Transl Behav Med 2020 Oct 08;10(4):1004-1015 [FREE Full text] [doi: 10.1093/tbm/ibz074] [Medline: 31120511]
26. Tombaugh TN. Trail making test A and B: normative data stratified by age and education. Arch Clin Neuropsychol 2004 Mar;19(2):203-214. [doi: 10.1016/S0887-6177(03)00039-8] [Medline: 15010086]
27. Blackburn HL, Benton AL. Revised administration and scoring of the digit span test. J Consult Psychol 1957 May;21(2):139-143. [doi: 10.1037/h0047235] [Medline: 13416432]
28. Schretlen D, Testa S, Pearlson G. Calibrated Neuropsychological Normative System. Lutz, FL: Psychological Assessment Resources; 2010.
29. Kaplan E, Goodglass H, Weintraub S. Boston naming test. Clin Neuropsychol 2016. [doi: 10.1037/t27208-000]
30. Salthouse TA. When does age-related cognitive decline begin? Neurobiol Aging 2009 May;30(4):507-514 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2008.09.023] [Medline: 19231028]
31. Benedict RH, Schretlen D, Groninger L, Brandt J. Hopkins verbal learning test – revised: normative data and analysis of inter-form and test-retest reliability. Clin Neuropsychol 2010 Aug 09;12(1):43-55. [doi: 10.1076/clin.12.1.43.1726]
32. Benedict RH, Schretlen D, Groninger L, Dobraski M, Shpritz B. Revision of the brief visuospatial memory test: studies of normal performance, reliability, and validity. Psychol Assess 1996 Jun;8(2):145-153. [doi: 10.1037/1040-3590.8.2.145]
33. All-in-one research solution for real-world data capture. Labfront. URL: https://www.labfront.com [accessed 2024-04-29]
34. Schretlen DJ, Winicki JM, Meyer SM, Testa SM, Pearlson GD, Gordon B. Development, psychometric properties, and validity of the hopkins adult reading test (HART). Clin Neuropsychol 2009 Aug;23(6):926-943. [doi: 10.1080/13854040802603684] [Medline: 19191072]
35. Bondi MW, Edmonds EC, Jak AJ, Clark LR, Delano-Wood L, McDonald CR, et al. Neuropsychological criteria for mild cognitive impairment improves diagnostic precision, biomarker associations, and progression rates. J Alzheimers Dis 2014;42(1):275-289 [FREE Full text] [doi: 10.3233/JAD-140276] [Medline: 24844687]
36. Farias ST, Mungas D, Reed BR, Cahn-Weiner D, Jagust W, Baynes K, et al. The measurement of everyday cognition (ECog): scale development and psychometric properties. Neuropsychology 2008 Jul;22(4):531-544 [FREE Full text] [doi: 10.1037/0894-4105.22.4.531] [Medline: 18590364]
37. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. NPJ Digit Med 2020;3:90 [FREE Full text] [doi: 10.1038/s41746-020-0297-4] [Medline: 32613085]
38. Chan A, Chan D, Lee H, Ng CC, Yeo AH. Reporting adherence, validity and physical activity measures of wearable activity trackers in medical research: a systematic review. Int J Med Inform 2022 May;160:104696 [FREE Full text] [doi: 10.1016/j.ijmedinf.2022.104696] [Medline: 35121356]
39. Tocci FL, Morey MC, Caves KM, Deberry J, Leahy GD, Hall K. Are older adults ready for wireless physical activity tracking devices? A comparison of commonly used tracking devices. J Am Geriatr Soc 2016 Jan;64(1):226-228. [doi: 10.1111/jgs.13895] [Medline: 26782889]

Abbreviations

AD: Alzheimer disease
ADRD: Alzheimer disease and related dementias
ECog: Everyday Cognition

EMA: ecological momentary assessment
MCI: mild cognitive impairment

Edited by D Gooch; submitted 11.12.24; peer-reviewed by Y Zhang, C Wu, NG Schaham; comments to author 04.02.25; revised version received 03.03.25; accepted 17.03.25; published 08.04.25.

Please cite as:
Holmqvist S, Kaplan M, Chaturvedi R, Shou H, Giovannetti T
Longitudinal and Combined Smartwatch and Ecological Momentary Assessment in Racially Diverse Older Adults: Feasibility, Adherence, and Acceptability Study
JMIR Hum Factors 2025;12:e69952
URL: https://humanfactors.jmir.org/2025/1/e69952
doi:10.2196/69952
PMID:

A Novel Just-in-Time Intervention for Promoting Safer Drinking Among College Students: App Testing Across 2 Independent Pre-Post Trials

Philip I Chow[1], PhD; Jessica Smith[1], MS; Ravjot Saini[2], MS; Christina Frederick[1], BSc; Connie Clark[2], BS; Maxwell Ritterband[2], BA; Jennifer P Halbert[1], MSPH; Kathryn Cheney[2], BA; Katharine E Daniel[1], PhD; Karen S Ingersoll[1], PhD

[1]School of Medicine, University of Virginia, 560 Ray C Hunt Dr, Charlottesville, VA, United States
[2]University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Philip I Chow, PhD
School of Medicine, University of Virginia, 560 Ray C Hunt Dr, Charlottesville, VA, United States

Abstract

Background: Binge drinking, which is linked to various immediate and long-term negative outcomes, is highly prevalent among US college students. Behavioral interventions delivered via mobile phones have a strong potential to help decrease the hazardous effects of binge drinking by promoting safer drinking behaviors.

Objective: This study aims to evaluate the preliminary efficacy of bhoos, a novel smartphone app designed to promote safer drinking behaviors among US college students. The app offers on-demand educational content about safer alcohol use, provides dynamic feedback as users log their alcohol consumption, and includes an interactive drink tracker that estimates blood alcohol content in real time.

Methods: The bhoos app was tested in 2 independent pre-post studies each lasting 4 weeks, among US college students aged 18 - 35 years. The primary outcome in both trials was students' self-reported confidence in using protective behavioral strategies related to drinking, with self-reported frequency of alcohol consumption over the past month examined as a secondary outcome.

Results: In study 1, bhoos was associated with increased confidence in using protective behavioral strategies. Students also endorsed the high usability of the app and reported acceptable levels of engagement. Study 2 replicated findings of increased confidence in using protective behavioral strategies, and demonstrated a reduction in the self-reported frequency of alcohol consumption.

Conclusions: Bhoos is a personalized, accessible, and highly scalable digital intervention with a strong potential to effectively address alcohol-related behaviors on college campuses.

(JMIR Hum Factors 2025;12:e69873) doi:10.2196/69873

KEYWORDS

alcohol; college students; smartphone intervention; binge drinking; safe drinking

Introduction

Overview

Binge drinking, defined as consuming more than 5 standard drinks for men or 4 for women within a 2-hour period [1,2], is highly prevalent among US college students. This is concerning because binge drinking is linked to various immediate and long-term negative outcomes, including lower academic performance, a higher incidence of sexual assault, drunk driving, motor vehicle accidents, organ damage, and premature death [2-5]. Additionally, research shows that college students consume more alcohol than their noncollege-attending peers [6,7], which is troubling given that excessive drinking during late adolescence and early adulthood is a strong predictor of alcohol use disorders later in life [8]. Therefore, taking proactive steps to address binge drinking among college students is critical to reducing both the short- and long-term consequences in this population.

Alcohol Interventions in US Colleges

A common approach adopted by US colleges and universities to address hazardous drinking among students involves programs that focus on changing attitudes, increasing knowledge, and modifying behaviors related to alcohol use. Programs like the Brief Alcohol Screening and Intervention for College Students [9,10] typically include brief motivational counseling sessions aimed at reducing students' positive alcohol expectancies, increasing their awareness of the consequences of drinking, and enhancing their readiness to change [11-15]. Despite the

widespread implementation of such alcohol education and intervention programs, national data show that binge drinking rates among college students have remained relatively stable for over a decade [5], highlighting the ongoing difficulty in effectively addressing this public health issue.

Delivering Alcohol Interventions Through Smartphone Apps

Safer drinking interventions delivered through smartphones may offer distinct advantages in addressing binge drinking among young adults due to their ubiquity, high use, constant presence, and engaging features. Among all US demographic groups, college students have some of the highest rates of smartphone ownership and use [16]. Capitalizing on this widespread adoption provides an opportunity to develop personalized interventions that can effectively engage students and help regulate their alcohol consumption. Compared with in-person interventions, smartphone app-based alcohol interventions are highly accessible, cost-effective, scalable for large student populations, and can offer personalized feedback based on users' behaviors [17-19].

Researchers have begun exploring the use of smartphone apps to deliver alcohol interventions both in the general population [18-20] and among college students specifically [21-25]. Reviews generally support the feasibility and potential of app-based interventions for addressing outcomes of problematic drinking in college students, though results have been inconsistent across studies [21,23]. One potential reason for this inconsistency may be the lack of sufficient tailoring or personalization for US college students, given that incorporating strategies like gamification to personalize digital interventions may increase user engagement and improve outcomes [26,27]. For example, providing users with dynamic feedback about their drinking, such as their estimated blood alcohol level, may help them to better regulate their alcohol consumption during a night out with their friends to avoid negative consequences. Furthermore, providing users with graphical displays of their current and past alcohol consumption patterns may help them understand how their drinking has impacted other aspects of their lives, such as their academic performance and personal health.

Our team conducted a formative study to gain a detailed understanding of US students' preferences for interventions and their patterns of smartphone use [28]. Based on our findings and through conducting iterative user-centered design testing, our team developed bhoos (see below for a detailed description of the app and its functions). Similar to other app interventions, bhoos provides on-demand psychoeducational content about safer alcohol use. A novel feature of bhoos, however, is its just-in-time approach. Through an interactive drink tracking feature, students can log their consumption of alcohol in real time, enabling the app to provide dynamic feedback on recommended safe drinking behaviors based on estimated real-time blood alcohol content [28]. The primary goal of bhoos is therefore to promote safer drinking behaviors, in large part by increasing students' confidence in engaging in protective strategies, with the potential secondary outcome of reducing overall alcohol consumption.

The Current Studies

We investigated the feasibility and preliminary impact of bhoos in 2 pre-post studies conducted among students at a mid-Atlantic university in the US over a 4-week period. The hypothesis for study 1 was that bhoos would lead to an increase in students' confidence in using protective behavioral strategies (ie, behaviors that are used while drinking to reduce alcohol use or limit alcohol-related problems) from baseline to postintervention. A secondary hypothesis was that bhoos would lead to a reduction in self-reported alcohol consumption from baseline to postintervention. Study 1 also evaluated the impact of small monetary incentives on drinking outcomes and how students use the app to log their drinks. As study 1 was the first to test bhoos in college students, we examined students' engagement with the app and their ratings of its usability. Focus groups were conducted at the end of study 1 to gather feedback on the app from participants. Study 2 was conducted as a replication of self-reported drinking outcomes in study 1 to determine the potential of testing bhoos in a future randomized trial.

Ethical Considerations

Both studies reported in the manuscript were approved by the University of Virginia Institutional Review Board for Social and Behavioral Sciences (IRB-SBS #4020, 5334). All participants provided informed consent. Data are deidentified. Participants in study 1 received gift cards via the web as compensation for completing the baseline (US $20) and postassessments (US $45). Participants randomized to receive added incentives in study 1 received an additional sum, up to US $30, based on their completion of 3 engagement milestones: US $10 for downloading the app, US $10 for logging 1 drink or dry day, and US $10 for logging a "streak" of 3 consecutive days of drinks or dry days. Participants in study 2 received gift cards via the web as compensation for completing the baseline (US $20) and postassessments (US $25).

Study 1

Methods

Study Overview and Participants

Study 1 consisted of 3 phases. The first phase evaluated the bhoos app in a simple pre-post trial design. The second phase randomized (1:1) participants to receive small additional monetary incentives or not for using the bhoos app in a second pre-post trial. The third phase consisted of focus groups to capture impressions of the app by a subset of participants from the second phase.

For all phases of study 1, college students were eligible to participate if they met the following criteria: (1) aged 18 to 35 years, (2) current enrollment verified via a university email address, and (3) currently owned and used a smartphone. Students were enrolled using traditional, online, and social media methods. Recruitment spanned June through September 2021. Recruitment materials instructed applicants to complete a web-based interest form including contact information, demographic information, and questions to determine initial

study eligibility. Research coordinators verified student status and identity with the university's internal people search. Eligible verified applicants were invited to participate, signed a web-based consent form, and were enrolled in the study.

The phase 1 group included 83 participants (mean age 20.8 years, SD 1.6 years; 68% self-identified as female; 58% self-identified as White, 16% self-identified as Asian or Hawaiian/Pacific Islander, 4% self-identified as Black, 12% self-identified as multiracial, 4% prefer not to answer, and 6% chose to leave the item blank), and phase 2 group included 172 participants (mean age 20.1 years, SD 1.8 years; 60% self-identified as female; 48% self-identified as White, 19% Asian or Hawaiian/Pacific Islander, 8% Black, 11% multiracial, 1% prefer not to answer, and 13% chose to leave the item blank). Assessments occurred at baseline and postintervention, 28 days later. All participants received gift cards via the web as compensation for completing the baseline (US $20) and postassessments (US $45). Participants in the phase 2 group who were randomized to not receive added incentives (n=86) were only provided compensation for completing the baseline and postassessments. However, those randomized to receive added incentives (n=86) received an additional sum, up to US $30, based on their completion of 3 engagement milestones: US $10 for downloading the app, US $10 for logging 1 drink or dry day, and US $10 for logging a "streak" of 3 consecutive days of drinks or dry days (see bhoos description, below).

To improve the app for future trials, the third phase of the trial included a subset of participants (n=18) from the phase 2 group who were invited to be part of debriefing focus groups that occurred 9 - 12 months after the initial pre-post trial. A total of 5 focus groups were conducted and qualitative information was collected regarding participants' general impressions of the app, what features of the app they liked most/least, and the degree to which the app helped them manage their drinking.

Intervention: Bhoos

Bhoos (pronounced [booz]) is an app whose name is a play on the word "booze," slang for alcohol, and the nickname for University of Virginia students, "hoos." The app provides on-demand educational content about safer alcohol use and dynamic feedback to users as they log their alcohol consumption in real time. Users can log the type of alcoholic drink they are consuming and each drink entry is time-stamped. The app provides real-time estimates of the user's blood alcohol concentration (estimated blood alcohol concentration) based on their self-identified sex, weight, and number and type of alcoholic drinks logged. In-app notifications are pushed to users based on their estimated blood alcohol concentrations, including information about recommended actions to stay safe and avoid overdrinking. Users can view their current and past drink history through a built-in dashboard. To encourage engagement with the app, users can establish a streak of logins whenever they log consecutive drinking days or dry days (ie, days in which users logged no alcoholic drinks). To encourage engagement, the app also allows users to rate activity level or sleep quality as a secondary health behavior. Materials about safe drinking tips and staying healthy can be accessed directly from the main page. Users can also learn about health-related resources in and around the university through the app on demand. The app design and content were informed by formative research and a think-aloud process with the target population prior to this trial [28].

Measures

Baseline Alcohol Use Severity

The 10-item Alcohol Use Disorders Identification Test (AUDIT) is the most widely used self-report measure of unhealthy alcohol use [29]. Scores range from 0 to 40, with higher scores indicating more unhealthy alcohol use. Prior studies on college students have suggested cutoff scores for low-risk drinkers (<7), hazardous drinkers (8-15), and alcohol-dependent drinkers (>15). Participants completed the AUDIT at baseline (Cronbach α=0.78) to characterize their alcohol use over the past year.

Alcohol Consumption in the Past Month

We used a modified 3-item measure based on the Daily Drinking Questionnaire [30] to assess students' self-reported drinking behavior. Students were initially asked "How often did you drink during the last month?" with response options ranging from 1 (did not drink at all) to 7 (once a day or more). Responses were analyzed as a single-item measure of the average number of days per week in the last month involving alcohol consumption, with higher scores indicating more days per week that involved alcohol consumption. Students who responded that they had drank at least once a month (ie, a score of ≥2) were then asked 2 follow-up questions: "Think of a typical weekend evening (Friday or Saturday) during the last month. How many standard drinks did you drink on that evening?" and "Think of the occasion (any day of the week) you drank the most during the last month. How many standard drinks did you drink?" Students responded to each of these questions by typing a number into an open field. Because these questions were intended to assess for average and maximum alcohol consumption, respectively, we analyzed responses to them as separate items.

Protective Behavioral Strategies

A modified version of the 20-item Protective Behavioral Strategies Scale [31-33] was administered at baseline (α=0.92) and postintervention (α=0.93) to assess students' confidence in using protective behavioral strategies, or behaviors adopted while drinking to limit their alcohol consumption or limit alcohol-related problems. Specifically, the instructions were modified from the original version to ask students to select the response option that "best fits your confidence level" for using various protective strategies when drinking. Participants responded to each item on a modified scale from 1 (extremely not confident) to 7 (extremely confident), with scores ranging from 20 to 140 [34,35]. Higher scores indicate more confidence in using protective behavioral strategies such as stopping or limiting alcohol consumption, changing the speed or frequency of drinking, and reducing the risk of serious harm by using a designated driver, going home with a friend, or protecting one's drink from adulterants.

App Usability

The 10-item System Usability Scale (SUS) [36] was administered postintervention to assess the usability of the bhoos app. The SUS provides a global view of subjective assessments of usability. We customized the items to specifically mention bhoos. Sample items include "I think that I would like to use bhoos frequently" and "I thought bhoos was easy to use." Responses to each item range from 1 (strongly disagree) to 5 (strongly agree). Possible scores on the SUS range from 0 to 100, with a higher score indicating higher overall usability of a system or program. The SUS has been used in roughly 3500 surveys within 273 studies evaluating a range of systems, interfaces, and programs [37]. Internal consistency of the SUS was good (α=0.84).

Plan for Analyses

To determine whether students' self-reported drinking outcomes differed across the 2 phases, including the incentive conditions in phase 2, 1-way ANOVAs were conducted using the phase 1, phase 2 incentivized, and phase 2 nonincentivized groups as within participants variables and difference scores for the drinking outcomes as the dependent variables. Difference scores in drinking outcomes between the baseline and postintervention time points served as the dependent variables. Next, for each of the phase 1 and phase 2 groups, paired *t* tests were conducted to obtain estimates of short-term changes in drinking and attitudinal measures from baseline to follow-up. Box-Cox transformations were performed for the drinking outcomes where there was evidence of nonnormality of data based on skewness (outside of −1 and +1) or kurtosis (outside of −2 and +2) being outside of the conventional acceptable ranges. The pattern of findings did not change when using transformed or raw variables. Analyses of self-reported drinking outcomes were performed using IBM SPSS Statistics (version 26.0.0). App use is described using descriptive statistics. Qualitative feedback from focus groups was reviewed by members of the study team (KI, CF, NA, and CC), and recurring themes were identified.

Results

Baseline Alcohol Use Severity

To characterize the sample, we examined AUDIT scores for all participants in study 1. Of the 234 students who completed at least the baseline measures, 133 (59%) were classified as low-risk drinkers (AUDIT Score ≤7), 93 (40%) were classified as hazardous drinkers (AUDIT Score between 8 and 15), and 8 (3%) were classified as alcohol dependent drinkers (AUDIT score >15).

Evaluating the Impact of Study Phases on Self-Reported Outcomes

There was no significant effect of the study phase or incentives on any of the self-reported drinking outcomes, for the average number of days per week in the last month involving alcohol consumption ($F_{2, 232}$=0.294, *P*=.75, η^2=.003), typical weekend evening drink consumption in the last month ($F_{2, 165}$=0.662, *P*=.52, η^2=.008), the maximum number of drinks consumed in the last month ($F_{2, 175}$=0.005, *P*=.99, η^2=.00), or protective behavioral strategies ($F_{2, 232}$=1.469, *P*=.23, η^2=.013). Table 1 contains all results of self-reported drinking outcomes, separately for each phase and incentive condition. Below, we report self-reported drinking-related outcomes overall for those in phases 1 and 2 combined, followed separately for each phase and incentive group.

Table . Self-reported drinking outcomes for study 1.

	Average number of days per week drinking in the last month, mean (SD)				Average number drinks per weekend in the last month, mean (SD)				Max number of drinks on one occasion in the last month, mean (SD)				Protective behaviors, mean (SD)			
	Pre	Post	*t* test (*df*)	*P* value	Pre	Post	*t* test (*df*)	*P* value	Pre	Post	*t* test (*df*)	*P* value	Pre	Post	*t* test (*df*)	*P* value
Phase 1	3.66 (1.31)	3.71 (1.34)	0.54 (75)	.59	3.64 (1.68)	3.61 (1.75)	0.11 (60)	.91	5.68 (3.16)	5.94 (3.12)	0.79 (65)	.43	98.20 (21.01)	100.30 (20.18)	0.98 (75)	.33
Phase 2 (no incentive)	2.99 (1.35)	3.01 (1.36)	0.22 (74)	.83	3.21 (2.06)	2.89 (1.95)	1.35 (50)	.18	4.53 (3.36)	4.59 (3.65)	0.27 (53)	.81	106.19 (24.74)	111.91 (23.28)	2.29 (76)	.03
Phase 2 (incentive)	3.05 (1.52)	2.99 (1.27)	0.55 (81)	.59	3.76 (1.92)	3.93 (2.75)	0.12 (52)	.91	5.93 (3.05)	5.98 (4.48)	0.64 (54)	.52	106.14 (26.05)	108.52 (24.56)	0.89 (80)	.38

Protective Behavioral Strategies

Overall, there was a significant increase in students' confidence in using protective behavioral strategies in drinking from baseline ($Mean_{pre}$ 103.58, SD 24.26) to postintervention ($Mean_{post}$ 106.97, SD 23.20; t_{233}=2.393, *P*=.02), such that students reported more confidence engaging in more protective behaviors while drinking to limit alcohol-related problems from drinking from baseline to postintervention.

As seen in Table 1, despite an increase in students' self-reported confidence in using protective behavioral strategies from baseline to postintervention for all 3 groups, the effect was not significant for students in the phase 1 group (t_{75}=0.98, *P*=.33), and the phase 2 incentive group (t_{80}=0.89, *P*=.38). There was,

however, a significant increase in confidence in using protective behavioral strategies for the phase 2 no incentive group (t_{76}=2.29, P=.03).

Alcohol Consumption in the Past Month

Overall, there was no significant change to the average number of days per week in the last month involving alcohol consumption, from baseline ($Mean_{pre}$ 3.23, SD 1.43) to postintervention ($Mean_{post}$ 3.25, SD 1.39; t_{232}=0.07, P=.95). There were also no significant changes to students' average number of drinks per weekend in the last month from baseline ($Mean_{pre}$ 3.52, SD 1.90) to postintervention ($Mean_{post}$ 3.44, SD 2.20; t_{164}=0.85, P=.40), nor were there significant changes to students' maximum number of drinks consumed in the last month from baseline ($Mean_{pre}$ 5.43, SD 3.46) to postintervention ($Mean_{post}$ 5.51, SD 3.83; t_{173}=0.20, P=.84).

As seen in Table 1, there were no significant changes to students' self-reported drinking frequency in the past month from baseline to postintervention for students in the phase 1 group (drinking frequency in the past month, t_{75}=0.54, P=.59; the average number of drinks per weekend in the last month, t_{60}=0.11, P=.91; the maximum number of drinks consumed in the last month, t_{65}=0.79, P=.43). This was also the case for the phase 2 no incentive group (drinking frequency in the past month, t_{74}=0.22, P=.83; the average number of drinks per weekend in the last month, t_{50}=1.35, P=.18; the maximum amount of drinks consumed in the last month, t_{53}=.27, P=.81), and the phase 2 incentive group (drinking frequency in the past month, t_{81}=0.55, P=.59; average number of drinks per weekend in the last month, t_{52}=0.12, P=.91; the maximum amount of drinks consumed in the last month, t_{54}=0.64, P=.52).

Usability of Bhoos

There was no significant effect of the study phase or incentives on self-reported usability of bhoos ($F_{2,\ 234}$=0.097, P=.908, η^2=.001). Usability scores for students in phase 1 ranged from 22.5 to 100 (mean 71.22, SD 17.21). Scores for students in the phase 2 nonincentive group ranged from 35 to 100 (mean 71.75, SD 16.96), and 37.50 to 100 (mean 72.38, SD 15.69) for those in the phase 2 incentive group. The average usability score (mean 71.80, SD 16.54) placed bhoos in the third quartile of all programs evaluated by the SUS [37].

Engagement With Bhoos

Nearly all participants (93.33%, 238 of 255) in study 1 downloaded the app. The majority of participants (85.49%, 218 of 255) logged at least 1 drink or dry day. Roughly two-thirds (160 of 255) of the participants logged at least 1 alcoholic drink, whereas nearly 3-quarters (189 of 255) logged at least 1 dry day. Slightly more than half (145 of 255) logged at least 1 streak. On average, participants used the app for 12.67 days (SD 9.96 d) of the 28 days of the intervention period.

Overall, providing added incentives did not seem to impact how students used the app to log their drinks. Among those in the phase 2 group, students randomized to receive incentives logged, on average, 11.16 drinks (SD 8.92) while those randomized to receive no added incentives logged, on average, 13.04 drinks (SD 16.62), t_{130}=0.924, and P=.357. Moreover, students randomized to receive incentives logged, on average, 15.50 dry days (SD 9.39) while those randomized to receive no added incentives logged, on average, 15.90 dry days (SD 8.81), t_{169}=0.282, and P=.78.

Qualitative Feedback From Focus Groups

The feedback obtained from focus groups revealed several common themes. Students found the bhoos app to be visually appealing and user-friendly. They expressed that the drinking chart within the app raised their awareness of their drinking habits over time, and they appreciated the ability to review their drinking history within the app. However, some students felt that the psychoeducational content within the app was too static, and they suggested that incorporating videos would be beneficial. Additionally, many students expressed a desire to track a broader range of secondary behaviors, including mood and stress.

In response to this feedback from the focus groups, we made several iterative improvements to the bhoos app in preparation for the next pilot study (study 2). Specifically, we enhanced the psychoeducational content by adding 4 short videos on relevant topics, including information on helping an intoxicated friend, knowing the signs of alcohol overdose, learning about alcohol tolerance, and understanding standard drink amounts. We also included more information on safety tips while drinking. We refined the app's functionality for users to track their mood or stress to identify patterns related to drinking events. Last, known bugs and glitches that students reported while using the app were addressed, enhancing bhoos' overall performance and reliability.

Discussion

Collectively, the results from study 1 provide preliminary support for the feasibility and usability of bhoos among college students in the United States. While the students in this sample did not report a reduction in drinking frequency, findings indicate that they were more confident in their ability to engage in protective behaviors while drinking, potentially enhancing their safety. Both app engagement data and mixed methods results confirm the usability of bhoos, while providing added monetary incentives did not seem to impact any of the observed outcomes or how students used bhoos to track their drinking. The qualitative feedback from students was instrumental in identifying areas for app improvement to enhance usability and engagement. Building on these findings, study 2 was conducted to examine whether the impact of bhoos on drinking-related outcomes would be replicated in an independent sample of college students.

Study 2

Methods

Study Overview and Participants

Eligibility criteria were identical to study 1. College students were recruited over the 2022 Fall semester using advertisements in dining halls. Although the initial goal was to enroll up to 200

students, the recruitment timeline was hampered by the continuation of the COVID-19 pandemic and a tragic event that occurred on university grounds on November 13th, 2022 (a shooting on University grounds that resulted in the murders of 3 students). The latest baseline questionnaires were completed before the event and enrollment of new participants was halted after the event. Participants who completed baseline questionnaires were permitted to complete the study and complete the postintervention questionnaires. In total, 43 students were recruited (60% students self-identified as women; 24 students self-identified as White, 12 students self-identified as Asian or Native Hawaiian, 2 students self-identified as Black, and 5 students self-identified as multiracial). Once again, assessments were administered at baseline and postintervention, 28 days later. All participants received gift cards via the web as compensation for completing the baseline (US $20) and postassessments (US $25). An added incentivization group was not included in study 2.

Measures and Plan for Analyses

The same self-report measures were administered in study 2 as in study 1. Specifically, baseline alcohol use severity was assessed by the 10-item AUDIT [29] (α=0.82). Alcohol consumption in the past month was assessed at baseline and postintervention with the 3-item measure used in study 1. Each item was treated as a separate single-item scale. The modified 20-item Protective Behavioral Strategies Scale [31-33] was administered at baseline (α=0.90) and postintervention (α=0.91) to assess students' confidence in using protective behavioral strategies. Finally, the usability of the bhoos app was assessed postintervention using the SUS [36] (α=0.75).

Paired *t* tests were used to investigate short-term changes in drinking and protective behaviors from baseline to follow-up. Due to the limited sample size, if the normality assumption was not met some analyses were computed again with the paired Wilcoxon Signed Rank Test. This did not change the interpretation of any of the results]. Analyses were performed using R (version 4.3.3; R Core Team). App use is described using descriptive statistics.

Results

Baseline Alcohol Use Severity

Of the 43 students, 25 (58%) were classified as low-risk drinkers (AUDIT Score ≤7), 11 (26%) were classified as hazardous drinkers (AUDIT Score between 8 and 15), and 6 (14%) were classified as alcohol dependent drinkers (AUDIT score >15).

Protective Behavioral Strategies

As expected, there was a significant increase in students' confidence in using protective behavioral strategies after using bhoos from baseline ($Mean_{pre}$ 107, SD 17.19) to postintervention ($Mean_{post}$ 113.37, SD 15.87; t_{31}=-2.79, *P*=.01), such that students reported confidence in engaging in more protective behaviors while drinking to limit alcohol-related problems from drinking after the intervention compared with before.

Alcohol Consumption in the Past Month

There was a significant decrease in the average number of days per week drinking alcohol in the last month from baseline ($Mean_{pre}$ 2.32, SD 1.27) to postintervention ($Mean_{post}$ 2.06, SD 1.15; t_{33}=2.32, *P*=.031). Among students who reported drinking, there was no significant change to their typical weekend evening drink consumption from baseline ($Mean_{pre}$ 3.83, SD 2.80) to postintervention ($Mean_{post}$ 3.87, SD 2.30; t_{29}=0, *P*=.92), nor was there a significant change to the maximum number of drinks they consumed in a single occasion the last month, from baseline ($Mean_{pre}$ 5.17, SD 4.07) to postintervention ($Mean_{post}$ 5.34, SD 4.50; t_{28}=0.24, *P*=.82).

Usability of Bhoos

Usability scores ranged from 55 to 98 (out of a possible 100). Overall, students reported higher usability of the revised version of bhoos (mean 77.93, SD 11.92) than in the original version tested in study 1. The average score placed bhoos in the upper 75% (fourth quartile) of all programs evaluated by the SUS [37].

Discussion

Collectively, the findings from study 2 reproduce the drinking-related results from study 1. Specifically, the findings of study 2 replicate the finding that bhoos is associated with an increase in confidence in using protective behaviors while drinking. It is worth noting that students in study 2 also reported reduced frequency of drinking, which may have resulted from improvements to the app from study 1, which is supported by our finding that students rated the app as more usable in study 2 than in study 1.

Overall Discussion

Principal Findings

Developing effective, accessible, and scalable interventions to address excessive drinking among college students is critical due to the significant public health concerns posed by alcohol misuse on college campuses. In 2 studies, we observed promising and consistent results regarding the impact and usability of the bhoos app, which effectively promoted confidence in using protective behaviors related to drinking among US college students.

The results suggest that bhoos' features, such as drink tracking and dynamic feedback, help to encourage responsible drinking behaviors, including monitoring alcohol intake, planning for safe transportation after drinking, or seeking help when needed. The ability to provide personalized, real-time feedback increases the relevance of the app to the individual needs and experiences of its users. By offering recommendations based on the user's drinking patterns and risk factors, bhoos may engage users more effectively than a generic one-size-fits-all program which enhances the likelihood of behavior change. Moreover, this more personalized approach to providing risk feedback aligns with the principles of Motivational Interviewing [38,39] and follows proven strategies for promoting behavior change [40-42]. Smartphone app-delivered interventions like bhoos offer several advantages over traditional approaches, including

the ability to deliver actionable feedback at critical moments. Additionally, they provide the potential for widespread accessibility and scalability. College students, in particular, maybe the most receptive to mobile apps for obtaining information and support for health-related decisions, given their high smartphone use [16]. The ease of dissemination and the low cost of smartphone apps make them a practical option for broad implementation on college campuses.

Limitations and Future Directions

The findings from our studies should be considered in light of several limitations. First, the reliance on infrequent and static measures of drinking behaviors introduces the possibility of recall bias. Moreover, the primary outcomes of both studies reflected students' confidence in using various protective behaviors while drinking and not the actual behaviors they engaged in. Future research could benefit from incorporating more ecologically valid approaches, such as ecological momentary assessment, perhaps administered within the app, to assess for actual protective behaviors related to drinking while limiting recall bias. Additionally, the pre-post design of our study limits our ability to evaluate the efficacy of bhoos, which could be better assessed through a randomized controlled trial. The short time frame between baseline and postassessment (4 weeks) restricts our understanding of the long-term sustainability of the app's effects. Future studies should explore the app's impact over longer periods, such as a full college semester. Moreover, a large percentage of students in both studies were White and female which may limit the generalizability of our findings to minorities and men. Finally, the tragic shooting that occurred on university grounds may have influenced students' drinking behaviors in study 2, although the extent of this impact is unclear. Future studies should examine how such events, and events more broadly (eg, political elections, performance of University athletic teams), affect student drinking behaviors.

General Conclusion

In summary, preliminary findings across 2 studies indicate that bhoos shows strong potential as a tool for promoting protective behaviors related to alcohol use among college students. While these studies help advance highly accessible and scalable approaches to address the consequences of binge drinking by US college students, further research is needed to fully understand the app's role in comprehensive alcohol harm reduction efforts within this population. Its personalized approach, combined with its accessibility and scalability, suggests that bhoos could be a valuable addition to the range of interventions available to address alcohol-related issues on campuses.

Acknowledgments

The authors thank the Board of Visitors and the President of the University of Virginia for funding this project through the Strategic Investment Fund. The authors also appreciate the support and guidance of the Executive Committee for the project "Reducing the Burden of Addictions in Virginia": Dr Robert Klesges, Dr Lee Ritterband, Dr Nassima Ait-Daoud Tiouririne, Dr Karen Ingersoll, Dr Christopher Holstege, and Dr Margaret Shupnik. Finally, the authors thank the undergraduate research assistants for this study: Grace Biegger, Aasritha Natarajan, David Vann, and Shruti Vachhani. We would also like to thank our development team, Gabe Heath, Nicole Le, and Brendan Murphey, for the design and development of the bhoos app intervention.

Conflicts of Interest

None delcared.

References

1. Goldman MS. Commentary on White, Kraus, and Swartzwelder (2006): "Many college freshmen drink at levels far beyond the binge threshold". Alcohol Clin Exp Res 2006 Jun;30(6):919-921. [doi: 10.1111/j.1530-0277.2006.00123.x] [Medline: 16737448]
2. White AM, Kraus CL, Swartzwelder HS. Many college freshmen drink at levels far beyond the binge threshold. Alcohol Clin Exp Res 2006 Jun;30(6):1006-1010. [doi: 10.1111/j.1530-0277.2006.00122.x] [Medline: 16737459]
3. Chassin L, Pitts SC, Prost J. Binge drinking trajectories from adolescence to emerging adulthood in a high-risk sample: predictors and substance abuse outcomes. J Consult Clin Psychol 2002 Feb;70(1):67-78. [Medline: 11860058]
4. Howland J, Rohsenow DJ, Greece JA, et al. The effects of binge drinking on college students' next-day academic test-taking performance and mood state. Addiction 2010 Apr;105(4):655-665. [doi: 10.1111/j.1360-0443.2009.02880.x] [Medline: 20403018]
5. Hingson R, Zha W, Smyth D. Magnitude and trends in heavy episodic drinking, alcohol-impaired driving, and alcohol-related mortality and overdose hospitalizations among emerging adults of college ages 18-24 in the United States, 1998-2014. J Stud Alcohol Drugs 2017 Jul;78(4):540-548. [doi: 10.15288/jsad.2017.78.540] [Medline: 28728636]
6. Blanco C, Okuda M, Wright C, et al. Mental health of college students and their non-college-attending peers: results from the national epidemiologic study on alcohol and related conditions. Arch Gen Psychiatry 2008 Dec;65(12):1429-1437. [doi: 10.1001/archpsyc.65.12.1429] [Medline: 19047530]
7. Slutske WS, Hunt-Carter EE, Nabors-Oberg RE, et al. Do college students drink more than their non-college-attending peers? evidence from a population-based longitudinal female twin study. J Abnorm Psychol 2004 Nov;113(4):530-540. [doi: 10.1037/0021-843X.113.4.530] [Medline: 15535786]

8. Huurre T, Lintonen T, Kaprio J, Pelkonen M, Marttunen M, Aro H. Adolescent risk factors for excessive alcohol use at age 32 years. a 16-year prospective follow-up study. Soc Psychiatry Psychiatr Epidemiol 2010 Jan;45(1):125-134. [doi: 10.1007/s00127-009-0048-y] [Medline: 19363578]
9. Dimeff LA, Baer JS, Kivlahan DR, Marlatt G. Brief Alcohol Screening and Intervention for College Students (BASICS): Guilford Press; 1999.
10. Fachini A, Aliane PP, Martinez EZ, Furtado EF. Efficacy of brief alcohol screening intervention for college students (BASICS): a meta-analysis of randomized controlled trials. Subst Abuse Treat Prev Policy 2012 Sep 12;7:40. [doi: 10.1186/1747-597X-7-40] [Medline: 22967716]
11. Dunn ME, Fried-Somerstein A, Flori JN, Hall TV, Dvorak RD. Reducing alcohol use in mandated college students: a comparison of a brief motivational intervention (BMI) and the expectancy challenge alcohol literacy curriculum (ECALC). Exp Clin Psychopharmacol 2020 Feb;28(1):87-98. [doi: 10.1037/pha0000290] [Medline: 31008641]
12. Gesualdo C, Pinquart M. Expectancy challenge interventions to reduce alcohol consumption among high school and college students: a meta-analysis. Psychol Addict Behav 2021 Nov;35(7):817-828. [doi: 10.1037/adb0000732] [Medline: 33856837]
13. Lewis BA, O'Neill HK. Alcohol expectancies and social deficits relating to problem drinking among college students. Addict Behav 2000;25(2):295-299. [doi: 10.1016/s0306-4603(99)00063-5] [Medline: 10795955]
14. Patrick ME, Maggs JL. Does drinking lead to sex? daily alcohol-sex behaviors and expectancies among college students. Psychol Addict Behav 2009 Sep;23(3):472-481. [doi: 10.1037/a0016097] [Medline: 19769431]
15. Rhew IC, Duckworth JC, Lee CM. The association between intended drinking contexts and alcohol expectancies in college students: a daily diary study. Addict Behav 2021 Sep;120:106967. [doi: 10.1016/j.addbeh.2021.106967] [Medline: 33971498]
16. Internet, broadband fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/?tabItem=d40cde3f-c455-4f0e-9be0-0aefcdaeee00 [accessed 2024-09-17]
17. Nahum-Shani I, Smith SN, Spring BJ, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
18. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone apps targeting alcohol and illicit substance use: systematic search in in commercial app stores and critical content analysis. JMIR Mhealth Uhealth 2019 Apr 22;7(4):e11831. [doi: 10.2196/11831] [Medline: 31008713]
19. Colbert S, Thornton L, Richmond R. Smartphone apps for managing alcohol consumption: a literature review. Addict Sci Clin Pract 2020 May 7;15(1):17. [doi: 10.1186/s13722-020-00190-x] [Medline: 32381062]
20. Hoeppner BB, Schick MR, Kelly LM, Hoeppner SS, Bergman B, Kelly JF. There is an app for that - or is there? a content analysis of publicly available smartphone apps for managing alcohol use. J Subst Abuse Treat 2017 Nov;82:67-73. [doi: 10.1016/j.jsat.2017.09.006] [Medline: 29021117]
21. Hutton A, Prichard I, Whitehead D, et al. mHealth interventions to reduce alcohol use in young people: a systematic review of the literature. Compr Child Adolesc Nurs 2020 Sep;43(3):171-202. [doi: 10.1080/24694193.2019.1616008] [Medline: 31192698]
22. Kazemi DM, Borsari B, Levine MJ, et al. Effectiveness of a theory-based mHealth intervention for high-risk drinking in college students. Subst Use Misuse 2020;55(10):1667-1676. [doi: 10.1080/10826084.2020.1756851] [Medline: 32394772]
23. Berman AH, Gajecki M, Sinadinovic K, Andersson C. Mobile interventions targeting risky drinking among university students: a review. Curr Addict Rep 2016;3(2):166-174. [doi: 10.1007/s40429-016-0099-6] [Medline: 27226948]
24. Riordan BC, Winter T, Carey KB, et al. A combined web based intervention and ecological momentary intervention for reducing alcohol use among incoming first-year university students: results from a three-arm randomised controlled trial. Addict Behav 2023 Jan;136:107471. [doi: 10.1016/j.addbeh.2022.107471] [Medline: 36081248]
25. Bertholet N, Schmutz E, Studer J, et al. Effect of a smartphone intervention as a secondary prevention for use among university students with unhealthy alcohol use: randomised controlled trial. BMJ 2023 Aug 16;382:e073713. [doi: 10.1136/bmj-2022-073713] [Medline: 37586742]
26. Floryan MR, Ritterband LM, Chow PI. Principles of gamification for Internet interventions. Transl Behav Med 2019 Nov 25;9(6):1131-1138. [doi: 10.1093/tbm/ibz041] [Medline: 30944933]
27. Floryan M, Chow PI, Schueller SM, Ritterband LM. The model of gamification principles for digital health interventions: evaluation of validity and potential utility. J Med Internet Res 2020 Jun 10;22(6):e16506. [doi: 10.2196/16506] [Medline: 32519965]
28. Smith JG, Alamiri NS, Biegger G, Frederick C, Halbert JP, Ingersoll KS. Think-aloud testing of a novel safer drinking app for college students during COVID-19: usability study. JMIR Form Res 2022 Feb 17;6(2):e32716. [doi: 10.2196/32716] [Medline: 35175210]
29. Babor TF, de la Fuente JR, Saunders J, Grant M. AUDIT, the alcohol use disorders identification test: guidelines for use in primary health care: World Health Organization; 1992:1-29.
30. Collins RL, Parks GA, Marlatt GA. Social determinants of alcohol consumption: the effects of social interaction and model status on the self-administration of alcohol. J Consult Clin Psychol 1985 Apr;53(2):189-200. [doi: 10.1037//0022-006x.53.2.189] [Medline: 3998247]

31. DeMartini KS, Palmer RS, Leeman RF, et al. Drinking less and drinking smarter: direct and indirect protective strategies in young adults. Psychol Addict Behav 2013 Sep;27(3):615-626. [doi: 10.1037/a0030475] [Medline: 23088406]
32. Martens MP, Ferrier AG, Cimini MD. Do protective behavioral strategies mediate the relationship between drinking motives and alcohol use in college students? J Stud Alcohol Drugs 2007 Jan;68(1):106-114. [doi: 10.15288/jsad.2007.68.106] [Medline: 17149524]
33. Sugarman DE, Carey KB. The relationship between drinking control strategies and college student alcohol use. Psychol Addict Behav 2007 Sep;21(3):338-345. [doi: 10.1037/0893-164X.21.3.338] [Medline: 17874884]
34. Madson MB, Arnau RC, Lambert SJ. Development and psychometric evaluation of the revised protective behavioral strategies scale. Psychol Assess 2013 Jun;25(2):556-567. [doi: 10.1037/a0031788] [Medline: 23397929]
35. Treloar H, Martens MP, McCarthy DM. The protective behavioral strategies scale-20: improved content validity of the serious harm reduction subscale. Psychol Assess 2015 Mar;27(1):340-346. [doi: 10.1037/pas0000071] [Medline: 25558969]
36. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, Colarado: Measuring Usability LLC; 2011.
37. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May 1;4(3):114-123 [FREE Full text]
38. Hettema J, Steele J, Miller WR. Motivational interviewing. Annu Rev Clin Psychol 2005;1(1):91-111. [doi: 10.1146/annurev.clinpsy.1.102803.143833] [Medline: 17716083]
39. Apodaca TR, Longabaugh R. Mechanisms of change in motivational interviewing: a review and preliminary evaluation of the evidence. Addiction 2009 May;104(5):705-715. [doi: 10.1111/j.1360-0443.2009.02527.x] [Medline: 19413785]
40. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
41. Bensley RJ, Mercer N, Brusk JJ, et al. The eHealth behavior management model: a stage-based approach to behavior change and management. Prev Chronic Dis 2004 Oct;1(4):A14. [Medline: 15670446]
42. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3. [doi: 10.2196/games.3139] [Medline: 25658754]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
SUS: System Usability Scale

Edited by A Kushniruk; submitted 10.12.24; peer-reviewed by RS Mohn, R Mertz; revised version received 18.02.25; accepted 04.03.25; published 10.04.25.

Please cite as:
Chow PI, Smith J, Saini R, Frederick C, Clark C, Ritterband M, Halbert JP, Cheney K, Daniel KE, Ingersoll KS
A Novel Just-in-Time Intervention for Promoting Safer Drinking Among College Students: App Testing Across 2 Independent Pre-Post Trials
JMIR Hum Factors 2025;12:e69873
URL: https://humanfactors.jmir.org/2025/1/e69873
doi:10.2196/69873

Comparison of an AI Chatbot With a Nurse Hotline in Reducing Anxiety and Depression Levels in the General Population: Pilot Randomized Controlled Trial

Chen Chen[1], MSc; Kok Tai Lam[1], MPhil; Ka Man Yip[1], MStat; Hung Kwan So[1], PhD; Terry Yat Sang Lum[2], PhD; Ian Chi Kei Wong[3,4,5], PhD; Jason C Yam[6], MD; Celine Sze Ling Chui[7], PhD; Patrick Ip[1], MD

[1]Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Room 115, 1/F, New Clinical Building, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong SAR, China (Hong Kong)

[2]Department of Social Work Administration, University of Hong Kong, Hong Kong SAR, China (Hong Kong)

[3]Department of Pharmacology and Pharmacy, Li Ka Shing Faculty of Medicine, University of Hong Kong, Hong Kong SAR, China (Hong Kong)

[4]School of Pharmacy, Medical Sciences Division, Macau University of Science and Technology, Macau, China

[5]School of Pharmacy, Aston University, Birmingham, United Kingdom

[6]Department of Ophthalmology and Visual Sciences, Chinese University of Hong Kong, Hong Kong SAR, China (Hong Kong)

[7]School of Nursing, University of Hong Kong, Hong Kong SAR, China (Hong Kong)

Corresponding Author:
Patrick Ip, MD
Department of Paediatrics and Adolescent Medicine, School of Clinical Medicine, Li Ka Shing Faculty of Medicine, University of Hong Kong, Room 115, 1/F, New Clinical Building, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong SAR, China (Hong Kong)

Abstract

Background: Artificial intelligence (AI) chatbots have been customized to deliver on-demand support for people with mental health problems. However, the effectiveness of AI chatbots in tackling mental health problems among the general public in Hong Kong remains unclear.

Objective: This study aimed to develop a local AI chatbot and compare the effectiveness of the AI chatbot with a conventional nurse hotline in reducing the level of anxiety and depression among individuals in Hong Kong.

Methods: This study was a pilot randomized controlled trial conducted from October 2022 to March 2023, involving 124 participants allocated randomly (1:1 ratio) into the AI chatbot and nurse hotline groups. Among these, 62 participants in the AI chatbot group and 41 in the nurse hotline group completed both the pre- and postquestionnaires, including the GAD-7 (Generalized Anxiety Disorder Scale-7), PHQ-9 (Patient Health Questionnaire-9), and satisfaction questionnaire. Comparisons were conducted using independent and paired sample *t* tests (2-tailed) and the χ^2 test to analyze changes in anxiety and depression levels.

Results: Compared to the mean baseline score of 5.13 (SD 4.623), the mean postdepression score in the chatbot group was 3.68 (SD 4.397), which was significantly lower (*P*=.008). Similarly, a reduced anxiety score was also observed after the chatbot test (pre vs post: mean 4.74, SD 4.742 vs mean 3.4, SD 3.748; *P*=.005), respectively. No significant differences were found in the pre-post scores for either depression (*P*=.38) or anxiety (*P*=.19). No statistically significant difference was observed in service satisfaction between the two platforms (*P*=.32).

Conclusions: The AI chatbot was comparable to the traditional nurse hotline in alleviating participants' anxiety and depression after responding to inquiries. Moreover, the AI chatbot has shown potential in alleviating short-term anxiety and depression compared to the nurse hotline. While the AI chatbot presents a promising solution for offering accessible strategies to the public, more extensive randomized controlled studies are necessary to further validate its effectiveness.

Trial Registration: ClinicalTrials.gov NCT06621134; https://clinicaltrials.gov/study/NCT06621134

(JMIR Hum Factors 2025;12:e65785) doi:10.2196/65785

KEYWORDS

AI chatbot; anxiety; depression; effectiveness; artificial intelligence

Introduction

An estimated 1 in 7 people in Hong Kong experience a common mental disorder in their lifetime, with anxiety and depression being the two most common mental health issues in Hong Kong, accounting for 19% and 14% among the generation population, respectively [1,2]. Nearly three-quarters of individuals experiencing a mental health disorder do not seek professional help for various reasons such as high cost of services, fear of judgment, or stigma [3]. To address these potential problems, artificial intelligence (AI) chatbots have been incorporated into digital interventions, particularly web- and smartphone-based apps, to enhance user experience and optimize individualized mental health [4,5]. Chatbots can provide a safe and confidential space for people to seek help for their mental health concerns. Additionally, they are available at any time throughout a study, making them a convenient option for those who may lack access to traditional mental health services or require support beyond regular business hours. Further efforts toward strengthening implementation of chatbots are needed, and their application to the general public should be explored to improve perceptions of general mental health and increase awareness of the importance of premedical interventions in the future [6].

Chatbot models have been developed in various countries and are widely used in psychology and clinical practice. For example, Tess, a psychological AI service, has been tailored to provide immediate assistance to caregiving specialists, individuals receiving care, and family caregivers within a nonprofit organization in the United States and Canada [7]. Among university students, chatbot applications have showed a significant decrease in anxiety symptoms [8]. Anna, an AI-driven chatbot, simulates the function of a therapist and offers gamified versions of evidence-based activities drawn from various therapeutic approaches. Participants who engaged in activities facilitated by Anna provided more detailed responses containing a higher number of positive relational terms [9]. A review of users' opinions, satisfaction, and attitudes regarding depression and anxiety chatbot apps found that users felt supported and confident while they used apps that were easy to navigate, affordable, and free of cost [10]. In Hong Kong, the first AI-driven Cantonese psychological support tool, the Pai.ACT mobile app, was developed for parents of children with special education needs [11]. However, there is currently a lack of evidence on the effectiveness of AI chatbots in tackling mental health problems in the general public.

Given the risk of anxiety and depression in Hong Kong and the limited availability of mental health services during the epidemic, this study aimed to compare the effectiveness of the developed AI chatbot with that of a conventional nurse hotline in alleviating psychological anxiety and depression in the general public. Additionally, this study also aimed to understand user satisfaction with chatbot usage and gather preliminary data for larger randomized controlled trials to advocate for the AI chatbot as a psychological support tool for individuals seeking information on the chatbot platform.

Methods

Recruitment

This study was designed as a randomized controlled trial (RCT) with two parallel groups. Participants were parents recruited through two school principal's networks. The inclusion criteria included (1) being able to use smartphones proficiently to interact with the AI chatbot, (2) being fluent in Chinese language used in the study to effectively communicate with the AI chatbot, and (3) providing an electronic consent form. Parents who were unwilling or unable to commit to the entire research process and had difficulties in reading and understanding Chinese were excluded. Invitation letters were sent to kindergarten and primary school principal groups from chief schools, and 124 parents responded to this study and provided an electronic consent form. Participants were then randomly allocated in a 1:1 ratio to either the AI chatbot or the nurse hotline group using block randomization with blocks of four. Participants immediately filled out the posttest questionnaires after communicating with the AI chatbot or nurse. Those who did not respond immediately received reminders until they completed the posttest questionnaires. Participants with risky mental health problems were followed up by phone calls and were encouraged to ask to connect with necessary medical services and mental health clinical departments.

Ethical Considerations

This study was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West (approval number: UW21-344). Informed consent was obtained from all participants in the study, ensuring that they were fully aware of the nature and possible consequences of their participation. Participants were informed that they had the right to opt out at any time without penalty. The original consent and approval covers secondary analysis without additional consent. Data collected from participants were deindentified to ensure privacy and confidentiality. Identifying information, including names, initials, or hospital numbers, was removed from all datasets to protect participant identities. Participation was voluntary, and individuals were informed that no financial or material incentives would be provided. Consolidated Standards of Reporting Trials (CONSORT) reporting guideline was adhered to in this study [12].

AI Chatbot Development Procedures

The development of the AI chatbot involved 7 key steps. First, the developer analyzed requirements by understanding common queries and focusing on providing accurate information to the public and assisting nurses. Relevant data was then collected from credible sources, and natural language processing (NLP) algorithms were used to train a robust model on diverse questions and answers. By integrating large language models, the chatbot could provide detailed responses. Next, the developer designed and developed the backend infrastructure, including databases, application programming interfaces, and integration with University of Hong Kong's website and WhatsApp. Extensive testing and validation ensured the chatbot's accuracy and user-friendliness, with a beta version tested by selected users. Finally, the chatbot was deployed on University of Hong

Kong's website and made accessible through WhatsApp, with a nurse hotline version developed to assist health care professionals in handling public inquiries.

The AI chatbot system comprises several essential components to provide accurate and relevant information. It begins with context identification, where the chatbot analyzes keywords and phrases to understand the user's query, ensuring accurate interpretation and relevant response. Next, the system conducts query analysis using NLP techniques to determine the user's intent and extract specific details such as dates, locations, and vaccination information. The query response system then leverages the trained NLP model and large language model–generated responses to provide accurate and contextually appropriate answers tailored to users' needs. Finally, the chatbot relies on a regularly updated and comprehensive dataset containing the latest information on COVID-19 vaccinations, government regulations, and health care procedures, ensuring the delivery of reliable and accurate responses.

Questionnaires

The Patient Health Questionnaire-9 (PHQ-9) is a self-assessment tool consisting of 9 questions that measure the frequency and severity of depressive symptoms over the last 2 weeks. Each question is derived from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*) criteria and is rated on a scale of 0 (not at all) to 3 (almost daily) [13]. The Generalized Anxiety Disorder Scale-7 (GAD-7) is a 7-item self-report scale that evaluates the frequency and severity of anxious thoughts and behaviors during the last 2 weeks. Items are based on the diagnostic criteria of the *DSM-IV* and scored from 0 (not at all) to 3 (nearly every day) [14]. The service satisfaction survey is a 3-item questionnaire for participants to report their degree of satisfaction after using the two platforms. The rating score ranges from 0 to 10, representing an increasing satisfaction from not at all likely to extremely likely. Based on the standardized cutoff that determines the severity of anxiety and depression, participants were divided into two groups: the no-risk group (total score<4) and the risk group (total score>4) [15].

Statistical Analysis

Quantitative variables were tested for normal distribution using the Kolmogorov-Smirnov test. Normally distributed continuous variables were expressed as mean (SD) and categorical variables were expressed as numbers and percentages. A per-protocol analysis was performed for all outcomes when comparing the 2 groups. The independent *t* test (2-tailed) was employed to compare the preanxiety or predepression, postanxiety or postdepression, pre-post difference scores, and service satisfaction between the AI chatbot and nurse hotline groups, respectively. The paired *t* test (2-tailed) was used to compare scores before and after group chat. Linear regression was conducted to analyze the difference in posttest scores between the AI chatbot and nurse hotline groups adjusted by pretest scores. Categorical variables were compared with the χ^2 test. All analyses were conducted using SPSS software (version 29.0; IBM Corp).

Results

From October 2022 to March 2023, all the participants (N=124) answered prequestionnaires, and 62 (62/62, 100%) participants in the AI chatbot group and 41 (41/62, 66.1%) in the nurse hotline group completed both the pre- and postquestionnaires (Figure 1). Comparisons within groups before and after the test are displayed in Table 1 and Figure 2. Compared to the mean baseline score of 5.13 (SD 4.623), the mean postdepression score 3.68 (SD 4.397) was significantly lower in the chatbot group (*P*=.008). Similarly, a reduction in anxiety score was also observed after using the chatbot test (pre vs post: mean 4.74, SD 4.742 vs mean 3.4, SD 3.748; *P*=.005).

Figure 1. Diagram of the study design. AI: artificial intelligence. PHQ-9: Patient Health Questionnaire-9; GAD-7: Generalized Anxiety Disorder Scale-7.

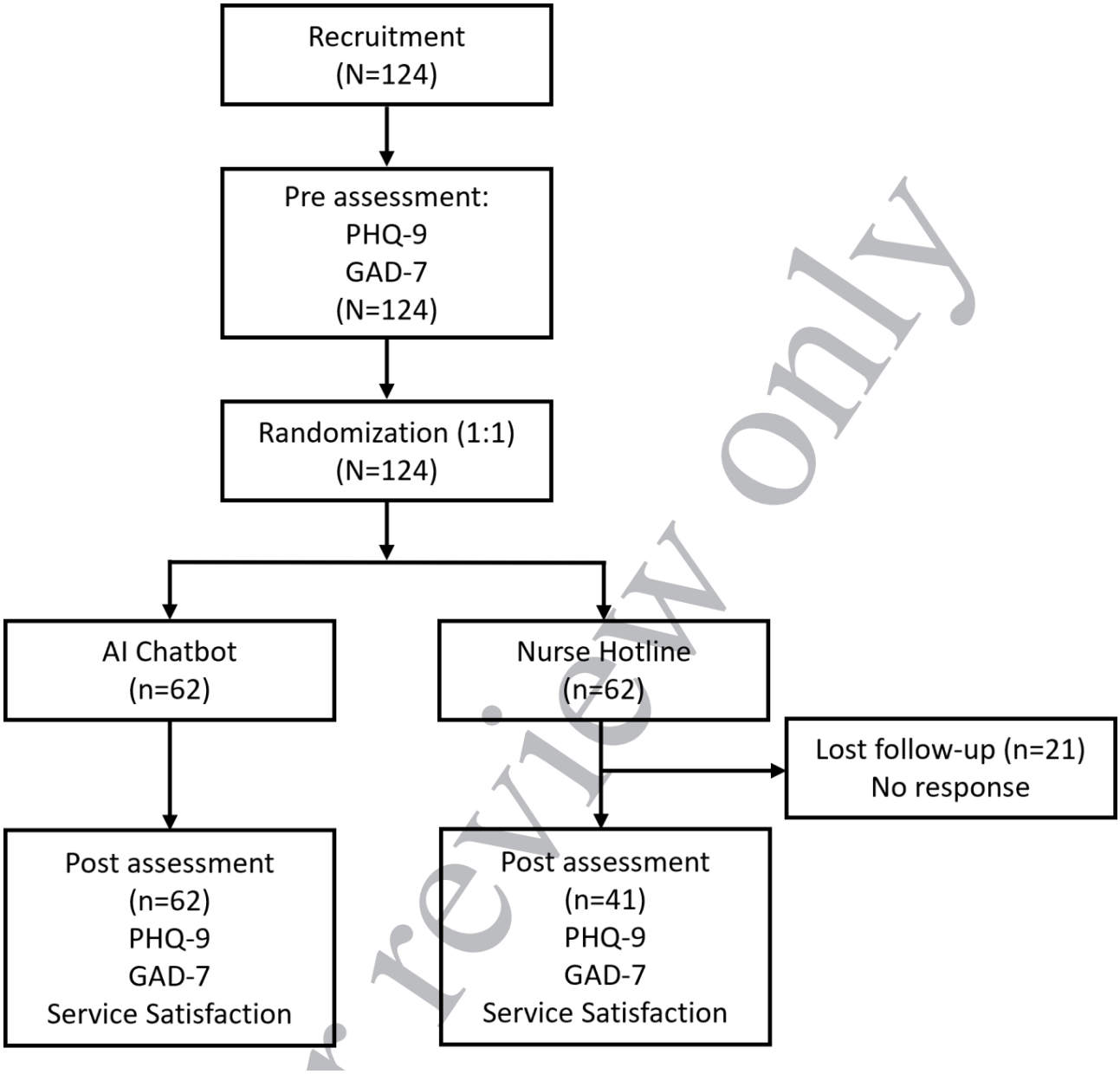


Table . Comparison of the anxiety and depression scores within groups before and after the test.

Questionnaires and groups (N=124)[a]	Pretest score, mean (SD)	Posttest score, mean (SD)	*P* value
PHQ-9[b]			
AI[c] chatbot (n=62)	5.13 (4.632)	3.68 (4.397)	.008
Nurse hotline (n=41)	5.46 (6.4)	4.76 (4.346)	.28
GAD-7[d]			
AI chatbot (n=62)	4.74 (4.742)	3.4 (3.748)	.005
Nurse hotline (n=41)	4.37 (5.439)	4.05 (4.455)	.63

[a]n=21 participants were lost to follow up.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]AI: artificial intelligence.

[d]GAD-7: Generalized Anxiety Disorder Scale-7.

Figure 2. Comparison of anxiety and depression scores within groups pre- and postintervention. GAD-7: Generalized Anxiety Disorder Scale-7; PHQ-9: Patient Health Questionnaire-9; ns: nonsignificant. **P<.05.

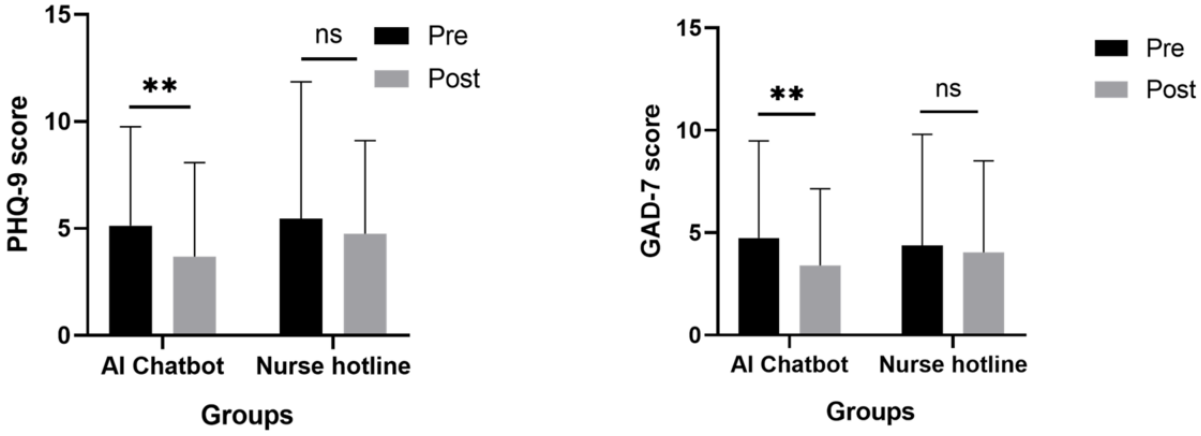


Table 2 shows there were no statistically significant differences at baseline between groups in depression (P=.76) and anxiety scores (P=.71). Although the postdepression and postanxiety scores in the chatbot group were slightly lower than those in the nurse hotline group, there was no statistical difference (all Ps>.05). No statistically significant difference was observed in service satisfaction between the 2 service platforms (P=.32). In the nurse hotline group, there were no notable differences in depression and anxiety scores before and after the test.

Table . Comparison of the before and after PHQ-9[a], GAD-7[b], and service satisfaction scores between the AI chatbot and nurse hotline groups.

Variables	AI chatbot scores (n=62), mean (SD)	Nurse hotline scores (n=41), mean (SD)	P value[c]
Pre–PHQ-9	5.13 (4.63)	5.46 (6.4)	.76
Post–PHQ-9	3.68 (4.40)	4.76 (4.35)	.43
Pre–GAD-7	4.74 (4.74)	4.37 (5.44)	.71
Post–GAD-7	3.40 (3.75)	4.05 (4.46)	.22
Service satisfaction	16.11 (6.88)	17.39 (5.59)	.32

[a] PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder Scale-7.

[c]Independent t test.

No significant variations were observed in the pre-post score differences for either depression (P=.38) or anxiety scores (P=.19) between the 2 groups (Table 3).

Table . Comparison of the pre-post anxiety and depression score differences between the artificial intelligence (AI) chatbot and nurse hotline groups.

Variables	Groups		
	AI chatbot (n=62), mean (SD)	Nurse hotline (n=41), mean (SD)	P value
Pre-post differences			
PHQ-9[a]	−1.452 (4.203)	−0.707 (4.161)	.38
GAD-7[b]	−1.339 (3.635)	−0.317 (4.15)	.19

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder Scale-7.

[c]Independent t-test was used to calculate P value.

The various test groups were not associated with the postdepression score (β=−.902, *P*=.18) or the postanxiety score (β=−.845, *P*=.17), after adjusting the prescores (Table 4).

Table . Regression analysis on the posttest score between the artificial intelligence chatbot and nurse hotline adjusted by pretest score (n=103).

Questionnaires and groups	β estimate	Standard error	*t* test (*df=100)*	*P* value
Post–PHQ-9[a] (ref: nurse group)				
Pre–PHQ-9	0.528	0.0615	8.591	<.001
Chatbot	−0.902	0.6720	−1.342	.18
Post–GAD-7[b] (ref: nurse group)				
Pre–GAD-7	0.531	0.0604	8.787	.001
Chatbot	−0.845	0.6148	−1.375	.17

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder Scale-7.

Based on the PHQ-9 and GAD-7 cutoff, the pre- and postscores were categorized into no-risk group and risk groups. No significant differences were observed in the prevalence of anxiety and depression before and after using the nurse hotline or AI chatbot (Table 5).

Table . Comparison of the anxiety and depression prevalence before and after using the nurse hotline and artificial intelligence (AI) chatbot among participants.

Groups	No-risk group (score≤4), n (%)	Risk group (score>4), n (%)	*P* value
Pre–PHQ-9[a]			
Nurse hotline (n=41)	27 (65.9)	14 (34.1)	.20
AI chatbot (n=62)	33 (53.2)	29 (46.8)	
Post–PHQ-9[a]			
Nurse hotline (n=41)	26 (63.4)	15 (36.6)	.33
AI chatbot (n=62)	45 (72.6)	17 (27.4)	
Pre–GAD-7[b]			
Nurse hotline	27 (65.9)	14 (34.1)	.34
AI chatbot	35 (56.5)	27 (43.5)	
Post–GAD-7[b]			
Nurse hotline	27 (65.9)	14 (34.1)	.47
AI chatbot	45 (72.6)	17 (27.4)	

[a]PHQ-9: Patient Health Questionnaire-9.

[b]GAD-7: Generalized Anxiety Disorder Scale-7.

Discussion

Principal Results

This study is among the first in Hong Kong to compare the effectiveness of AI chatbots and nursing hotlines in addressing mental health–related inquiries. The AI chatbot proved effective in reducing anxiety and depression during the COVID-19 outbreak and was comparable to the conventional nursing hotline. It suggests the potential use of AI chatbots as a complementary approach in mental health interventions, particularly when AI chatbots can provide timelier and uninterrupted 24-hour support to users in need during the challenging periods such as a pandemic.

Previous research comparing the effectiveness of the AI chatbot and humans in improving mental health is controversial. Our study found that both the AI chatbot and nurse hotline had similar effects in alleviating depression and anxiety. The results are consistent with another study comparing the text-only chatbot and voice-based digital human interactions as responders to mental health questions, which reported that the text-only chatbot was more user-friendly, although there were no significant differences in electroencephalography measurements [16]. No differences were observed in another RCT comparing the effects of AI chatbot (ie, ChatGPT-3.5) and nurse education on preoperative anxiety reduction among groups [17]. However, one study comparing physician and AI chatbot responses to patient questions posted on a public social media forum found that the chatbot generated higher quality and empathetic

responses to patient questions than physicians [18]. Additionally, the follow up process took less time for the nurse hotline group compared to the chatbot group. This was primarily because participants in the nurse hotline group were quicker in responding to follow up questionnaires while the chatbot group typically required more than one attempt to gather responses. It is feasible to use the AI chatbots to provide instant information to users about mental health knowledge and basic strategies to dissipate their temporary anxiety and depressed mood. However, the AI chatbot can only be a facilitator, and humans will not be replaced in the long-term professional treatment. The sample size in the above-mentioned studies is very small, which may have limited the power of statistical analysis. Therefore, more RCTs with larger sample sizes are necessary to confirm differences between the AI chatbot and human-based services.

Our study showed a significant reduction in the levels of anxiety and depression after using the AI chatbot. This finding is supported by a study from Argentina, which showed a substantial decrease in anxiety symptoms in university students using *Tess* [8]. Another study reported that the use of the Elomia chatbot for 4 weeks resulted in a significant decrease in the symptoms of anxiety and depression along with reduced negative effects [19]. AI chatbots can offer information to users in a quick and easy format. They can also be programmed to answer specific questions about a certain condition, such as what to do during a medical crisis or what to expect during a medical procedure. The AI chatbot could lower the barriers to care by helping patients access help more quickly and efficiently [20]. AI chatbots appear to have a positive impact on depression and anxiety in a wide range of populations, including university students [8], short-course treatment patients [21], and preoperative patients [17]. The potential benefit of chatbot–assisted mental health support is that it can provide helpful information about depressive moods, especially for users who have difficulty in communicating emotions to other humans [22]. Besides, the chatbot offers real-time relief, emotional support, and instant messages on basic knowledge about health-related questions. AI chatbots can help reduce anxiety and depression by providing accessible, anonymous, consistent, and immediate emotional support. They serve as a valuable supplement to conventional therapy and mental health care, offering users a safe space to express their feelings and find effective coping strategies.

Our study differs from previous ones by addressing a research gap in the broader population. Earlier studies primarily concentrated on specific groups, such as college students aged 18-33 [8], younger individuals aged 19-23 [19], and older adults aged 60 years [17]. The sample size in our study (N=124) aligns with past studies, which typically included approximately 100 participants. A study with a larger sample size of 412 younger participants [19] also found AI chatbots to be effective in alleviating anxiety and depression symptoms. Notably, our study was conducted during the COVID-19 pandemic, unlike previous studies, which were conducted under normal circumstances. Hence, additional research is necessary to understand the efficacy and impact of AI chatbots across different age groups, demographics, and environmental contexts.

Our study was carried out during the COVID-19 outbreak, revealing that AI chatbot interventions are versatile and can be applied in various situations, such as during future pandemics or for disseminating information to prevent other diseases. There is great potential in AI chatbot designing and application during an epidemic like COVID-19 [23]; AI chatbots not only help in reducing depression and anxiety at any given time but also can be embedded into the health care systems, to help patients describe their symptoms and provide preliminary guidance on whether they need to seek medical attention [24]. This can assist in triaging patients based on the severity of their symptoms. They can help patients book appointments with health care providers, reducing the burden on administrative staff. AI Chatbots can send reminders to patients to take their medications on time, improving medication adherence [24]. They can provide ongoing emotional support and coping strategies for individuals dealing with mental health issues like anxiety, depression, and stress. This can complement traditional therapy and make mental health care more accessible. Chatbots can also deliver personalized health education to patients, helping them understand their conditions, treatment options, and preventive measures.

The findings of this study comparing AI chatbots and nurse hotlines in alleviating the levels of anxiety and depression during the COVID-19 pandemic can be generalized to a certain extent, given that the general population was the target group. This broad target population ensures that the results are applicable to a wide range of individuals who may have experienced similar mental health challenges during the pandemic. However, the generalizability may be limited by factors such as the specific characteristics of the sample used in the study, including their demographics and the severity of their anxiety and depression symptoms. Additionally, the dynamic nature of AI chatbots, which can be influenced by various factors such as the phrasing of questions and ongoing optimization processes, may also affect the generalizability of the results. The study's design and methodology provide a solid foundation for replicability. The use of an RCT framework, which is a robust design for evaluating interventions, increases the likelihood that similar studies can be conducted with comparable results. The clear documentation of the study's procedures, including the recruitment process, data collection methods, and statistical analysis techniques facilitates replication by other researchers. Furthermore, the study's focus on a common and significant issue—mental health during a pandemic—ensures that there is a continued interest and relevance in conducting similar studies. However, replicability may be challenged by the need for access to similar populations and resources, as well as the evolving nature of AI chatbot technology, which may require adaptation of the intervention for future studies.

Although AI chatbots have a potential complementary function in mental health and health-related fields, certain drawbacks must be considered. For example, the safety and privacy concerns remain unclear [25]. Chatbots are unable to deliver the emotional support and personal bond that a certified mental health expert may provide. They cannot deliver a diagnosis or management strategy for mental health disorders. Therefore, it is crucial to use AI chatbots as a complement to, rather than a

substitute for, specialized mental health services. Moreover, further refinement to the AI chatbot is necessary for establishing an integrated system for recognition-alert-reporting system for mental health problems. After early detection and identification of depression and anxiety, the AI chatbot should be designed to automatically send self-regulation advice to the users to handle their negative emotions as well as send reminder messages to encourage them to seek medical services. Additionally, the question coverage of the AI chatbot is limited; it covers common mental health–related questions that may overlook the scope of other health problems. The scope of the population using chatbots should be expanded among the public. Further studies are required to draw solid conclusions about the effectiveness and safety of chatbots.

Limitations

One of the limitations of our study is the small sample size. Toward the end of the COVID-19 outbreak, there was a notable lack of enthusiasm among parents to participate in this study. Only two school principal groups were involved, which may have introduced a selection bias. Additionally, the study period was too short to recruit sufficient participants; however, the preliminary results showed a significant reduction in anxiety and depression after using the AI chatbot, supporting the applicability of using AI chatbots to relieve negative emotions during the next epidemic crisis. Demographic data and other confounding factors were not collected due to privacy concerns; these factors will be included in the following larger randomized control study.

Conclusions

The AI chatbot is as effective as the nurse hotline in answering mental health–related questions. Besides, the AI chatbot can reduce short-term anxiety and depression levels compared to the nurse hotline. The developed AI chatbot shows promise as a complementary tool to provide available intervention strategies to the public, particularly for families with limited access to mental health services. This study provides insights into the potential application of using AI chatbots to provide immediate information to the public and mitigate public emotional distress during future epidemic emergencies.

Acknowledgments

This study was funded by the Collaborative Research Fund, University Grants Committee (reference no. C7149-20GF) and General Research Fund, University Grants Committee (reference no. 17606523). All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Authors' Contributions

Conceptualization: PI, TYSL, ICKW, CSLC
Data curation: KLT
Formal analysis: KMY, CC
Funding acquisition: PI
Investigation: KLT
Methodology: TYSL, ICKW, CSLC
Project administration: PI
Supervision: PI
Writing—original draft: CC
Writing—review & editing: CC, HKS, KMY

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH checklist (V1.6.1).
[PDF File, 89 KB - humanfactors_v12i1e65785_app1.pdf]

References

1. Mental health review report. Food and Health Bureau, Hong Kong Special Administrative Region Government. 2017. URL: https://www.healthbureau.gov.hk/download/press_and_publications/otherinfo/180500_mhr/e_mhr_full_report.pdf [accessed 2024-07-15]
2. Choi EPH, Hui BPH, Wan EYF. Depression and Anxiety in Hong Kong during COVID-19. Int J Environ Res Public Health 2020 May 25;17(10):3740. [doi: 10.3390/ijerph17103740] [Medline: 32466251]
3. Lam LCW, Wong CSM, Wang MJ, et al. Prevalence, psychosocial correlates and service utilization of depressive and anxiety disorders in Hong Kong: the Hong Kong Mental Morbidity Survey (HKMMS). Soc Psychiatry Psychiatr Epidemiol 2015 Sep;50(9):1379-1388. [doi: 10.1007/s00127-015-1014-5] [Medline: 25660760]

4. D'Alfonso S. AI in mental health. Curr Opin Psychol 2020 Dec;36:112-117. [doi: 10.1016/j.copsyc.2020.04.005] [Medline: 32604065]
5. Tso WWY, Leung LK, Chow MSC, et al. Mental health of children with special educational needs and the return to in-person learning after the COVID-19 pandemic. JAMA Netw Open 2023 Dec 1;6(12):e2346106. [doi: 10.1001/jamanetworkopen.2023.46106] [Medline: 38060228]
6. Torous J, Bucci S, Bell IH, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
7. Joerin A, Rauws M, Ackerman ML. Psychological artificial intelligence service, Tess: delivering on-demand support to patients and their caregivers: technical report. Cureus 2019 Jan 28;11(1):e3972. [doi: 10.7759/cureus.3972] [Medline: 30956924]
8. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial intelligence-based chatbot for anxiety and depression in university students: pilot randomized controlled trial. JMIR Form Res 2021 Aug 12;5(8):e20678. [doi: 10.2196/20678] [Medline: 34092548]
9. Boucher EM, Harake NR, Ward HE, et al. Artificially intelligent chatbots in digital mental health interventions: a review. Expert Rev Med Devices 2021 Dec;18(sup1):37-49. [doi: 10.1080/17434440.2021.2013200] [Medline: 34872429]
10. Ahmed A, Aziz S, Khalifa M, et al. Thematic analysis on user reviews for depression and anxiety chatbot apps: machine learning approach. JMIR Form Res 2022 Mar 11;6(3):e27654. [doi: 10.2196/27654] [Medline: 35275069]
11. CUHK nursing develops mobile app "pai.ACT" the first AI-driven Cantonese psychological support tool for parents of SEN children in Hong Kong. Faculty of Medicine, The Chinese University of Hong Kong. 2023. URL: https://www.med.cuhk.edu.hk/press-releases/cuhk-nursing-develops-mobile-app-pai-act-the-first-ai-driven-cantonese-psychological-support-tool-for-parents-of-sen-children-in-hong-kong [accessed 2024-07-15]
12. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. EQUATOR Network. URL: https://www.equator-network.org/reporting-guidelines/consort [accessed 2024-07-15]
13. Yu X, Tam WWS, Wong PTK, Lam TH, Stewart SM. The Patient Health Questionnaire-9 for measuring depressive symptoms among the general population in Hong Kong. Compr Psychiatry 2012 Jan;53(1):95-102. [doi: 10.1016/j.comppsych.2010.11.002] [Medline: 21193179]
14. Robinson CM, Klenck SC, Norton PJ. Psychometric properties of the Generalized Anxiety Disorder Questionnaire for DSM-IV among four racial groups. Cogn Behav Ther 2010;39(4):251-261. [doi: 10.1080/16506073.2010.486841] [Medline: 20830629]
15. Mental health screening: Patient Health Questionnaire-9 (PHQ-9). National HIV Curriculum. URL: https://www.hiv.uw.edu/page/mental-health-screening/phq-9 [accessed 2024-07-15]
16. Thunström AO, Carlsen HK, Ali L, Larson T, Hellström A, Steingrimsson S. Usability comparison among healthy participants of an anthropomorphic digital human and a text-based chatbot as a responder to questions on mental health: randomized controlled trial. JMIR Hum Factors 2024 Apr 29;11:e54581. [doi: 10.2196/54581] [Medline: 38683664]
17. Yahagi M, Hiruta R, Miyauchi C, Tanaka S, Taguchi A, Yaguchi Y. Comparison of conventional anesthesia nurse education and an artificial intelligence chatbot (ChatGPT) intervention on preoperative anxiety: a randomized controlled trial. J Perianesth Nurs 2024 Oct;39(5):767-771. [doi: 10.1016/j.jopan.2023.12.005] [Medline: 38520470]
18. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 1;183(6):589-596. [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
19. Romanovskyi O, Pidbutska N, Knysh A. Elomia chatbot: the effectiveness of artificial intelligence in the fight for mental health. In: Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Systems (COLINS 2021): Computational Linguistics and Intelligent Systems; 2021:1215-1224 URL: https://api.semanticscholar.org/CorpusID:235271819 [accessed 2024-07-17]
20. Habicht J, Viswanathan S, Carrington B, Hauser TU, Harper R, Rollwage M. Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. Nat Med 2024 Feb;30(2):595-602. [doi: 10.1038/s41591-023-02766-x] [Medline: 38317020]
21. Zhong W, Luo J, Zhang H. The therapeutic effectiveness of artificial intelligence-based chatbots in alleviation of depressive and anxiety symptoms in short-course treatments: a systematic review and meta-analysis. J Affect Disord 2024 Jul 1;356:459-469. [doi: 10.1016/j.jad.2024.04.057] [Medline: 38631422]
22. Chin H, Song H, Baek G, et al. The potential of chatbots for emotional support and promoting mental well-being in different cultures: mixed methods study. J Med Internet Res 2023 Oct 20;25:e51712. [doi: 10.2196/51712] [Medline: 37862063]
23. Battineni G, Chintalapudi NO, Amenta F. AI chatbot design during an epidemic like the novel coronavirus. Healthcare (Basel) 2020 Jun 3;8(2):154. [doi: 10.3390/healthcare8020154] [Medline: 32503298]
24. Lee HO, Kang JO, Yeo JO. Medical specialty recommendations by an artificial intelligence chatbot on a smartphone: development and deployment. J Med Internet Res 2021 May 6;23(5):e27460. [doi: 10.2196/27460] [Medline: 33882012]

25. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Household M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021. [doi: 10.2196/16021] [Medline: 32673216]

Abbreviations

AI: artificial intelligence
CONSORT: Consolidated Standards of Reporting Trials
***DSM-IV*:** *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*
GAD-7: Generalized Anxiety Disorder Scale-7
NLP: natural language processing
PHQ-9: Patient Health Questionnaire-9
RCT: randomized controlled trial

Edited by A Kushniruk; submitted 26.08.24; peer-reviewed by S Steingrimsson, Y Yang; revised version received 14.01.25; accepted 15.01.25; published 06.03.25.

Please cite as:
Chen C, Lam KT, Yip KM, So HK, Lum TYS, Wong ICK, Yam JC, Chui CSL, Ip P
Comparison of an AI Chatbot With a Nurse Hotline in Reducing Anxiety and Depression Levels in the General Population: Pilot Randomized Controlled Trial
JMIR Hum Factors 2025;12:e65785
URL: https://humanfactors.jmir.org/2025/1/e65785
doi:10.2196/65785

Original Paper

Development of Digital Strategies for Reducing Sedentary Behavior in a Hybrid Office Environment: Modified Delphi Study

Iris Parés-Salomón[1,2*], MSc; Cristina Vaqué-Crusellas[1,3*], PhD; Alan Coffey[4*], MSc; Bette Loef[5*], PhD; Karin I Proper[5,6*], Prof Dr; Anna M Señé-Mir[7*], PhD; Anna Puig-Ribera[2*], PhD; Kieran P Dowd[4*], PhD; Judit Bort-Roig[1,2*], PhD

[1]Faculty of Health Sciences and Welfare, University of Vic – Central University of Catalonia (UVic-UCC), Vic, Spain

[2]Sport and Physical Activity Research Group, Institute for Research and Innovation in Life and Health Sciences in Central Catalonia (IRIS-CC), University of Vic – Central University of Catalonia (UVic-UCC), Vic, Spain

[3]Research group on Methodology, Methods, Models and Outcomes of Health and Social Sciences (M3O), Institute for Research and Innovation in Life Sciences and Health in Central Catalonia (IRIS-CC), Vic, Spain

[4]SHE Research Centre, Department of Sport and Health Sciences, Technological University of the Shannon, Athlone, County Westmeath, Ireland

[5]Centre for Prevention, Lifestyle and Health, National Institute for Public Health and the Environment, Bilthoven, The Netherlands

[6]Department of Public and Occupational Health, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

[7]Sport and Physical Activity Research Group, Sport and Physical Activity Studies Centre, University of Vic-Central University of Catalonia, Vic, Spain

[*]all authors contributed equally

Corresponding Author:
Iris Parés-Salomón, MSc
Faculty of Health Sciences and Welfare
University of Vic – Central University of Catalonia (UVic-UCC)
C Sagrada Família, 7
Vic, 08500
Spain
Phone: 34 938 815 50
Email: iris.pares@uvic.cat

Abstract

Background: Hybrid work is the new modus operandi for many office workers, leading to more sedentary behavior than office-only working. Given the potential of digital interventions to reduce sedentary behavior and the current lack of studies evaluating these interventions for home office settings, it is crucial to develop digital interventions for such contexts involving all stakeholders.

Objective: This study aimed to reach expert consensus on the most feasible work strategies and the most usable digital elements as a delivery method to reduce sedentary behavior in the home office context.

Methods: A modified Delphi study including 3 survey rounds and focus groups was conducted to achieve consensus. The first Delphi round consisted of two 9-point Likert scales for assessing the feasibility of work strategies and the potential usefulness of digital elements to deliver the strategies. The work strategies were identified and selected from a scoping review, a systematic review, and 2 qualitative studies involving managers and employees. The median and mean absolute deviation from the median for each item are reported. The second round involved 2 ranking lists with the highly feasible strategies and highly useful digital elements based on round 1 responses to order the list according to experts' preferences. The weighted average ranking for each item was calculated to determine the most highly ranked work strategies and digital elements. The third round encompassed work strategies with a weight above the median from round 2 to be matched with the most useful digital elements to implement each strategy. In total, 4 focus groups were additionally conducted to gain a greater understanding of the findings from the Delphi phase. Focus groups were analyzed using the principles of reflexive thematic analysis.

Results: A total of 27 international experts in the field of occupational health participated in the first round, with response rates of 86% (25/29) and 66% (19/29) in rounds 2 and 3, respectively, and 52% (15/29) in the focus groups. Consensus was achieved on 18 work strategies and 16 digital elements. Feedback on activity progress and goal achievement; creating an action plan; and standing while reading, answering phone calls, or conducting videoconferences were the most feasible work strategies, whereas wrist-based activity trackers, a combination of media, and app interfaces in smartphones were the most useful digital elements.

Moreover, experts highlighted the requirement of combining multiple levels of strategies, such as social support, physical environment, and individual strategies, to enhance their implementation and effectiveness in reducing sedentary behavior when working from home.

Conclusions: This expert consensus provided a foundation for developing digital interventions for sedentary behavior in home office workers. Ongoing interventions should enable the evaluation of feasible strategies delivered via useful digital elements in home office or hybrid contexts.

(JMIR Hum Factors 2025;12:e59405) doi:10.2196/59405

KEYWORDS

sedentary behavior; office work; home office; hybrid work; technology; Delphi

Introduction

Background

Workplace sedentary behavior (SB) is associated with poorer health and work-related outcomes, such as musculoskeletal disorders, reduced mental and occupational well-being, job dissatisfaction, and fatigue [1-4]. According to the World Health Organization, the workplace is one of the main settings to reduce SB in the adult population [5], with office workers, who spend an average of 77% of their working day sitting [6], being a key target group [7]. After the COVID-19 pandemic, hybrid work, alternating work in the office and at home, has become the new work paradigm for many desk-based jobs [8-10]. Hybrid work results in even higher levels of SB than working solely in the office [11], increasing the risk of musculoskeletal health problems, mental health issues, and reduced work productivity [11,12].

Developing strategies to reduce and break up workplace SB, replacing it with activities other than sitting, such as walking or standing, has been the focus of recent studies [13,14]. In particular, the use of digital strategies in desk-based jobs (eg, delivered via mobile phone, activity tracker, or desktop computer) has the potential to provide scheduled prompts, deliver information, give automated tailored feedback, and mediate organizational support and social influences [15]. Recent evidence has highlighted the capability of digital interventions to reduce workplace SB and its associated harmful effects on health and work-related outcomes among office workers [16-18]. However, this research has focused on the traditional office environment, with a dearth of research available on the use of digital interventions to reduce of employees' SB while working remotely or undertaking hybrid work [19,20].

Multicomponent digital interventions that combine environmental and organizational changes and provision of information and counseling have been reported as more effective than single-component interventions in reducing occupational SB among office workers [16,21]. In addition, the grounding of interventions in a theory of behavior change is key in the success and behavior maintenance of multicomponent interventions [22-24]. Given the limited evidence base on the effectiveness of interventions targeting hybrid work, it is crucial to develop multicomponent digital interventions grounded in sound theory of behavior change to reduce occupational SB among home office workers.

Objectives

To contribute to a successful implementation of an intervention and, thereby, to its effectiveness, involving stakeholders such as researchers, experts in the addressed topic, and the target population in the development of interventions is necessary to identify needs, priorities, and potential solutions [25]. A previous qualitative study identified which factors influenced employees' ability to reduce SB when working from home from the employers' perspective, whereas additional qualitative evidence identified the factors influencing the reduction in SB from the employees' perspective [26]. This study aimed to go beyond the existing research gaps to incorporate expert opinions to gain a complete picture of all stakeholders and reach consensus on the feasibility of digital work strategies to reduce SB in a home office context. Therefore, the aim of this study was to reach expert consensus on the most feasible and usable digital work strategies as a delivery method to reduce SB in a home office context.

Methods

Design

A modified Delphi study was used to elicit a consensus from a spectrum of experts in the field of occupational health. It is a flexible approach combining quantitative and qualitative data. Qualitative data can be collected before, after, or between Delphi rounds [27,28]. This study encompassed 3 survey rounds and subsequent web-based focus group sessions. This study is part of the initial modeling of a multicomponent intervention and of a larger European project (ie, the Click2Move program) aimed at tackling the challenge of occupational sedentarism in hybrid office workers.

This study was based on the recommendations of the Conducting and Reporting Delphi Studies guidelines [29] and the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [30].

Ethical Considerations

This project was approved by the Research Ethics Committee of the University of Vic–Central University of Catalonia (250/2023). This study was conducted in accordance with relevant guidelines and regulations. The invitation email contained general information about the study and a link with further details of the participant information statement, including assurances of anonymity and confidentiality and emphasizing that participation was completely voluntary and participants

could withdraw at any time without any consequences. In compliance with ethical research practices, digital informed consent was obtained from all participants before initiating the survey. No incentives were provided to the participants for their involvement.

Participants and Recruitment

International experts (N=57) identified as leading or established researchers in the field of occupational health (with a specific focus on modification of physical activity [PA] behavior in the workplace), well-being, or computer- or desk-based jobs specialists identified by the research team were invited via email to participate. The experts were selected through a search of published articles that identified highly active authors in the field, whereas the job specialists were identified from professional networks of the involved institutions. The exclusion criterion was not being able to understand and communicate in English.

To maximize participation in the Delphi survey, a reminder email was sent within 1 week. After completing the 3 Delphi rounds, participants who signed the informed consent form for the Delphi study (29/57, 51%) were invited via email to participate in a focus group.

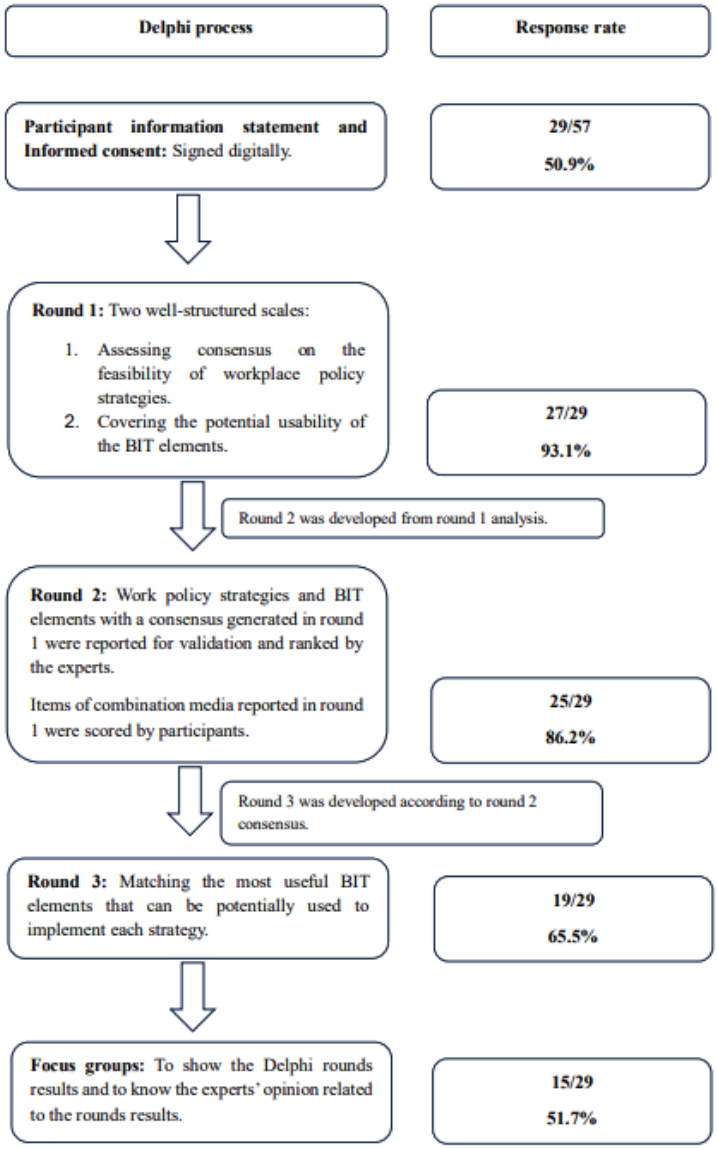
Procedure

Delphi Survey

Overview

In total, 3 survey rounds were conducted using the LimeSurvey software (LimeSurvey GmbH; Multimedia Appendix 1), which were piloted by 4 members of the research team to ensure the appropriateness of the questionnaire items, wording, and functionality and identify any issues or gaps. No changes in wording were made to the Delphi surveys. If participants agreed to take part, they received an automatic email with the link to the first survey. Participants had 10 days to complete each round of the survey. Individual responses were anonymized, and all participants were invited to complete the second and third rounds of the survey regardless of their participation status in the previous rounds. Attrition was prevented through a short period between rounds and by providing feedback on study findings. A summary of the entire Delphi process is provided in Figure 1.

Figure 1. Delphi process. BIT: behavioral intervention technology.



Round 1

Round 1 was carried out between January 13 and 23, 2023. Participants were provided with an explanation of the survey and were asked to provide demographic information (ie, age, gender, and employer organization). Two 9-point Likert scales were used.

The first scale assessed participants' opinions on the feasibility of work strategies (n=36) for reducing SB when working from home. The work strategies had been selected from a scoping review of the vast majority of real-life PA strategies implemented [31], a systematic review of effective digital interventions to reduce SB [21], and 2 qualitative studies that further explored the perceptions of managers and employees on the factors influencing the ability of employees to reduce their SB in a home office context [26]. The strategies were classified based on the Behavior Change Wheel (ie, environmental planning and service provision, 9/36, 25% of the strategies; regulation, guidelines, and restriction, 15/36, 42% of the strategies; and communication and social support, 12/36, 33% of the strategies) [32].

Participants were invited to score each work strategy. Scores ranging between 1 and 3 suggested a lack of feasibility, scores between 4 and 6 suggested neutral feasibility, and scores between 7 and 9 suggested high feasibility. This scoring was based on the 9-point Grading of Recommendations Assessment, Development, and Evaluation methodology [33]. The feasibility scores considered the 4 dimensions of feasibility: social validity or acceptability, integration into the existing system, practicality, and adaptability (Table 1) [34].

The second scale covered the potential usefulness of the technology-based behavior change elements according to the behavioral intervention technology (BIT) model (ie, messaging and social support [n=6 elements], push notifications [n=5 elements], information delivery [n=9 elements], report and visualization [n=4 elements], digital log [n=3 elements], and passive data collection [n=7 elements]) [35]. Scoring was based on the 9-point Grading of Recommendations Assessment,

Development, and Evaluation scale divided into 3 categories: not useful (1-3), neutral (4-6), and highly useful (7-9) [33]. Participants were also invited to provide additional items or observations through an open-ended question for each strategy or BIT element category.

Table 1. Feasibility dimensions for the strategies and the related research question for each dimension.

Dimension	Research question
Social validity or acceptability	Is the strategy appropriate, reasonable, fair, and potentially effective for home-based work?
Integration into the existing system	To what extent are the strategies aligned with the infrastructure of home office settings?
Practicality	Can the initiative be implemented using the available resources, time, training, and materials in home-work offices?
Adaptability	Is the initiative flexible enough to fit across the diverse needs of home office work?

Round 2

Round 2 started on February 6, 2023, and finished on February 16, 2023. Work strategies from round 1 were ranked based on the *highly feasible* scores for work strategies and *highly useful* scores for BIT elements. A total of 35 work strategies (rather than 36) were ranked in round 2 based on recommendations from round 1 (a similar approach was implemented for work strategies and BIT elements). In this second round, participants were asked to place each work strategy in order of preference considering the most feasible strategies to reduce SB when working from home. The same procedure was followed for the BIT elements. Additional items that appeared in the first survey were scored based on the round 1 approach (ie, the 9-point Likert scale).

Round 3

Round 3 was launched on February 17, 2023, and could be answered until February 27, 2023. According to the preferred order of the items ranked in round 2, participants were asked to match the most useful BIT elements that could potentially be used to implement each of the work strategies to reduce SB when working from home. The work strategies included in this round were those with a weight above the median.

Focus Groups

To explore in more depth the experts' opinions and understand their perspectives, focus groups took place.

Focus groups were conducted in English or Spanish by a researcher (AC or IP-S) and a moderator (IP-S, JB-R, or AMS-M) with previous experience facilitating focus groups. The focus groups were held on the web using Microsoft Teams in April 2023. The focus groups lasted a total of 60 minutes. The size of the groups was small (4 participants on average) due to scheduling difficulties and time zone differences among countries.

A guide was developed based on the 3 Delphi rounds (Multimedia Appendix 2). However, the researchers did not follow it in full detail, using the freedom to explore thoughts and ideas based on the responses from each group. First, there was a brief welcome and introduction to explain the aim of the focus group, obtain consent for the recording of the session, and introduce each participant. The results of every Delphi round were then shown. Subsequently, experts were asked about what factors they thought made some strategies more feasible than others considering the 4 dimensions (outlined in Table 1) and what factors may enhance participation among the BIT elements. The last element of the focus group covered the need for any adaptation before the implementation of the strategies in the hybrid work model. The focus groups were recorded, transcribed verbatim, and translated into English if necessary to be analyzed later.

Data Management and Analysis

Study uptake was calculated as the proportion of participants who accessed the round 1 Delphi survey to that of all potential participants identified. Retention rate referred to the proportion of completed responses for each round and participation in the focus groups. In round 1, responses to the Likert scale in the *highly feasible* or *highly useful* category (score from 7 to 9) were calculated. The median and mean absolute deviation from the median (MAD-M) for each item were also reported to assess the strength and extent of agreement. Medians of 7 to 9 were defined as strong support, medians of 4 to 6 were defined as moderate support, and medians of 1 to 3 were defined as no support. The MAD-M describes the variation in the median values [36]. In round 2, we calculated the weighted average ranking for each item to determine the most highly ranked work strategies and BIT elements for reducing SB among home office workers. The respondents' most preferred choice had the largest weight, and the least preferred choice had a weight of 1. For achieving consensus, the order-weighted average process was followed through the aggregation of the strategies evaluated in the previous round to be validated in round 2 [37]. Items that achieved consensus were those that were ranked with a weighted average above the median. The additional items that emerged from round 1 were assessed through a 9-point Likert scale in round 2, and the median and MAD-M were calculated. In round 3, the matched items for each work strategy and BIT element were reported, as well as the percentage of respondents.

The focus group analysis was conducted using the principles of reflexive thematic analysis [38,39]. First, a researcher (IPS) familiarized themselves with the data, reading and rereading carefully each transcript to obtain a general impression of the main opinion of the experts and identify repeated concepts. Subsequently, according to the subjectivity of the researcher, an initial open coding process was generated from interesting features and relevant data searching for evidence for themes, namely, patterns of shared meaning underpinned by a central organizing concept. Reflexive thematic analysis was approached

in an inductive, semantic, and realistic way. In other words, the construction of codes and themes was directed by the content of the data but not by a topic summary. Codes, concepts, and the core category were recoded and relabeled if necessary after each step until the codes provided the intended depth of the insights. The analytic outputs obtained from the codes and through the active engagement of the researcher with their data were then conceptualized in potential themes and subthemes. Finally, the themes were reviewed, defined, and named. At the end of each step, data were shared with an experienced supervisor (CV-C) for reflecting on how the data were initially coded and consider things that might have been overlooked. MAXQDA 2022 (VERBI GmbH) was used for the analysis.

To ensure a robust analytical process, we considered key aspects of quality and rigor. The first aspect was reflexivity, which involves critical reflection of researcher perspectives and its integration within the analysis and interpretation of the data to develop themes [40]. The study team comprised researchers with expertise in PA and SB in the workplace who have experienced working from home since before or after COVID-19 pandemic. IP-S, who carried out the analysis, is a Spanish physiotherapist and osteopath PhD student who has been working from home for 3 years; CV-C, the supervisor during the analysis, is an experienced nutritionist with a background in health promotion and qualitative research who has been working from home since the COVID-19 pandemic. Second, confirmability, the objectivity of the findings [41], was attained through a critical friend approach whereby an experienced supervisor reviewed interpretations and introduced alternative perspectives, increasing objectivity and confirming the accuracy of the findings. Third, dependability, the reliability and consistency of the study [41], was attained through detailed research procedures (ie, rigorous data collection techniques and well-documented procedures and analyses). Credibility was then built on the reflexivity and triangulation of the data, incorporating more than one researcher in the analysis process [42]. Furthermore, a critical friend approach was used whereby an experienced qualitative researcher, CV-C, critically analyzed the preliminary themes and ensured that the selected quotes provided support for the themes developed. Finally, transferability was enhanced through the transparent and detailed descriptions of the participants and research process [42].

Results

Participant Characteristics

A total of 51% (29/57) of the invited experts expressed their interest in participating in the study through the digital informed consent form. Of these 29 participants, 27 (93%) completed the first round, 25 (86%) completed the second round, and 19 (66%) completed the third round. Of those 29 participants, 15 (52%) joined the focus groups.

Most of the first-round Delphi respondents and focus group participants identified as women (16/27, 59% and 8/15, 53%, respectively), with 48% (13/27) of first-round respondents aged between 36 and 45 years.

A wide variety of countries across Europe were represented. However, non-European countries such as Australia (5/29, 17% of the participants) were also represented.

Delphi Survey

Feasibility of Work Strategies

Round 1

Multimedia Appendix 3 shows the results of round 1, which considered the 36 work strategies divided into 3 categories. A total of 69% (25/36) of the work strategies were ranked as highly feasible.

In total, 7 environmental planning and service provision work strategies were rated as highly feasible with strong support. Activity trackers (20/27, 74%; median 8, MAD-M 1.41); workstation accessories (19/27, 70%; median 8, MAD-M 1.59); and relocation of home office supplies such as the bin, printer, or water source (19/27, 70%; median 8, MAD-M 1.30) were the 3 work strategies considered as highly feasible for reducing SB when working from home. Active workstations equipped with treadmills were reported to be unfeasible (19/27, 70%; median 2, MAD-M 1.48).

In total, 53% (8/15) of the work strategies regarding guidelines, regulations, and restrictions were rated as highly feasible with strong support. Self-monitoring of sedentary and activity behaviors (20/27, 74%; median 8, MAD-M 1.30), substitution of sedentary tasks (eg, reading, answering a phone call, or participating in a videoconference) with standing or incidental movements (20/27, 74%; median 7, MAD-M 1.37), and creating an action plan (20/27, 74%; median 7, MAD-M 1.11) were the most highly useful. The least feasible strategy was the locking or blocking of the screen or keyboard for specific durations (11/26, 42%; median 5, MAD-M 2.19).

Regarding the communication and social support work strategies, feedback on activity progress and achievement of goals (22/27, 81%; median 7, MAD-M 1.74); providing information to increase awareness and knowledge (20/27, 74%; median 8, MAD-M 1.37); and the provision of awards, rewards, or incentives by the company to achieve goals (19/27, 70%; median 7, MAD-M 1) were ranked as the most feasible for reducing SB when working from home. Most of the work strategies included in this category (10/12, 83%) were rated by the respondents as highly feasible, and none of them were considered unfeasible. No new items emerged from open-ended questions.

Round 2

The ranking list of the work strategies with the weighted average (range 14.6-0.48) is reported in Table 2. Most items reported as highly feasible when working from home in round 1 (18/25, 72%) were scored above the median (3.64) in round 2, with the exception of self-selected reminders (3.24); active breaks such as stretching, walking, or strengthening exercises (2.36); real-time records (2.32); high or height-adjustable chairs (1.96); exercise accessories (1.52); and social networking for sharing experiences (0.48).

Table 2. Ranking of highly feasible work strategies (results of round 2).

Strategy	Weighted average score
Feedback on activity progress and goal achievement	14.6
Create an action plan—increase standing breaks or replace sitting time, how long, how often, when, and how (eg, when the phone rings)	13.04
Standing while reading, answering phone calls, or conducting videoconferences	13.04
Providing information to increase awareness and knowledge of the dangers associated with prolonged sedentary behavior and the potential benefits of reducing it or breaking it up	12.88
Self-monitoring sedentary and activity behaviors (ie, activity tracker or a diary log)	10.76
Short breaks (5-10 min) approximately every 60 min of sitting time	10.36
Scheduling (blocking) 5-10–min breaks between meetings on the calendar	8.6
Information and support about the strategies and goals and reminders	8.04
Setting tailored goals for reducing sitting time	7.24
Activity demonstrations to perform during the breaks	7.12
Height-adjustable desks or desk platforms	6.92
Relocation of home office supplies (eg, bins and printers)	6.8
Incidental moving while reading, answering phone calls, or conducting videoconferences	5.76
Awards, rewards, or incentives to achieve goals or recommendations	5.48
Standing desks	4.6
Point-of-choice or point-of-decision prompts	4.4
Workstation accessories (seated footrests, standing footrests, or sit-stand antifatigue mats for standing desks)	4.24
Motivational messages from managers	3.64
Wellness coaches supporting the employees during breaks	3.48
Self-selected reminders to achieve goals	3.24
Blocking the screen or keyboard for breaking up sitting time unless conducting a meeting	3.24
Active workstation with underdesk stepper or pedaling device	2.52
Short breaks (5-10 min) approximately every 30 min of sitting time	2.48
Team or individual activity challenges	2.4
Active breaks, such as stretching, walking, or performing strengthening exercises for ≥10 min	2.36
Scheduling (blocking) snack breaks on the calendar	2.32
Real-time records or feedback	2.32
Active lunchtime, such as Pilates, yoga, walking, or cycling	2.04
High chairs or height-adjustable chairs	1.96
Social comparison	1.68
Exercise accessories, such as rubber bands, wooden sticks, or mats	1.52
Active workstation equipped with treadmill	1.44
Short breaks (5-10 min) approximately every 40 min of sitting time	1.36
Competition among peers	0.92
Social networking for sharing experiences	0.48

Usefulness of BIT Elements

Round 1

Multimedia Appendix 4 shows the results of round 1 considering the usefulness of the BIT elements divided into 5 categories.

In total, 33% (2/6) of the BIT elements from the messaging and social support category were ranked as highly useful with strong support: social challenge features (20/27, 74%; median 7, MAD-M 1.22) and gamification (14/26, 54%; median 7, MAD-M 1.78). None of the messaging and social support items were considered not useful. Smartphone app and desktop application interfaces were identified to be the 2 most useful

BIT elements within the push notification category (12/27, 44%; median 7; and MAD-M 1.41 for smartphone app interface and 15/27, 56%; median 7; and MAD-M 1.18 for desktop application interface). The other 3 push notification elements were ranked as highly useful or neutral with medians between 5 and 6, indicating moderate support. A combination of different media was the preferred item in the delivery of information category (14/24, 58%; median 7, MAD-M 1.71), with the other elements within the category receiving moderate support and being considered neutrally or highly useful. A total of 15 new combined digital items, such as videos and images, moving images on websites, or reminders combined with social challenges, emerged from open-ended responses to the combination of media element in the information delivery category. Blending was considered more useful than isolated media. Real-time data (19/27, 70%; median 7, MAD-M 1.48) and data summaries via the app interface (14/27, 52%; median 7, MAD-M 1.67) were considered the most useful in the report and visualization category. The 2 remaining items (data summaries via email or SMS text message and websites) were ranked as neutral. The 3 most useful BIT elements to track activity were wrist-based activity trackers (17/26, 65%; median 7, MAD-M 1.69) and using a mobile phone diary (17/27, 63%; median 7, MAD-M 1.41) or smartphone sensors (13/26, 50%; median 7, MAD-M 1.61). The only element considered as not useful to track activity was waist-based activity trackers (10/26, 38%; median 5, MAD-M 1.85).

Round 2

Overall, BIT elements reported as highly useful to support work strategies were ranked at the top in each category (Table 3). In the messaging and social support category, items ranked above the median score (2.52) coincided with highly useful items in round 1 (ie, social challenge features and gamification), but chats (eg, WhatsApp groups) were added to the top of the ranking. SMS text messages were added to the 2 BIT elements in the push notification category considered as highly useful in round 1, which were scored above the median (2.00) in round 2. In addition to combination of media, videos, interfaces in smartphone apps and desktop applications (same as in the push notification category), images, and websites were among the items scored above the median (3.24) in the information delivery category. The new items that emerged from the open-ended responses in round 1, specifying potential combinations of media for information delivery described as highly useful, were a combination of mobile phone apps and desktop applications; a combination of personalized messages, reminders, and social challenges; a combination of videos, text, and images within an interactive website; and a combination of a mobile app with social media (Table 4). In the report and visualization category, the 2 items ranked as highly useful in round 1 remained above the median (2.32) in round 2. The median of the BIT elements to track activity was 3.56. In round 2, a total of 2 more items were added as highly useful to the 3 identified in round 1: leg-based activity tracker and computer software. In total, 17 BIT elements were considered highly useful in round 2.

Table 3. Ranking of highly usable behavioral intervention technology (BIT) elements (results of round 2).

BIT element to communicate information	Weighted average scores (SD)
Messaging and social support	
Social challenge features (cooperative activities, eg, reach 10,000 steps among all members of the department)	5.16
Gamification features (competition among colleagues)	4.48
Chats (eg, WhatsApp groups)	3.28
Emails	1.76
Calls	1.4
Forums	1.64
Push notifications	
App interface—text and sound and vibration in smartphones	4
Application interface—computer screen notification in desktop computers	3.08
SMS text messages	2
Chats	1.64
Emails	1.52
Information delivery	
Combination of media	7.6
Videos	6
App interface—smartphones	4.44
Images	4.36
Application interface—desktops	3.24
Websites	3.24
Emails	3.16
SMS text messages	2.36
Audios	1.44
Report and visualization	
Real-time data via app interface	2.96
Data summary via app interface	2.84
Data summary via email or SMS text message	1.8
Data summary via website	1.24
BIT elements to track activity	
Mobile phone diary (manual entry)	7.72
Web-based questionnaire (manual entry)	3.52
Computer software diary (manual entry)	1.6
Passive data collection	
Wrist-based activity tracker	8.6
Leg-based activity tracker	3.64
Smartphone sensors	3.6
Computer software	3.6
Waist-based activity tracker	3.28
Electronic workstation	3.08
Cushions on chairs	2.04

Table 4. Results for the new items from round 1 scored in round 2—behavioral intervention technology (BIT) elements.

New BIT element	Level of usefulness, n (%)									Median (MAD-M[a])
	Not useful			Neutrally useful			Highly useful			
	Score of 1	Score of 2	Score of 3	Score of 4	Score of 5	Score of 6	Score of 7	Score of 8	Score of 9	
Mixture of videos, information, and images within an interactive website (n=23)	0 (0)	1 (4)	2 (9)	2 (9)	0 (0)	6 (26)	5 (22)	6 (26)	1 (4)	6 (1.48)
App and social media (n=23)	0 (0)	1 (4)	3 (13)	0 (0)	2 (9)	1 (4)	10 (43)	4 (17)	2 (9)	7 (1.30)
Videos and images (n=23)	0 (0)	1 (4)	1 (4)	1 (4)	4 (17)	3 (13)	12 (52)	1 (4)	0 (0)	6 (1.17)
Audio and video combination with an explanation (educational purposes; n=23)	0 (0)	0 (0)	2 (9)	1 (4)	4 (17)	6 (26)	2 (9)	3 (13)	3 (13)	6 (1.48)
Moving images on a website (n=22)	1 (5)	1 (5)	3 (14)	2 (9)	4 (18)	7 (32)	3 (14)	1 (5)	0 (0)	5 (1.41)
Smartphone and desktop with website or app for background information (n=22)	1 (5)	0 (0)	1 (5)	0 (0)	3 (14)	3 (14)	6 (27)	7 (32)	1 (5)	6.5 (1.39)
Personalized messages and app (n=22)	0 (0)	0 (0)	0 (0)	0 (0)	1 (5)	3 (14)	8 (36)	5 (23)	5 (23)	7 (0.91)
Reminders combined with an app (n=22)	0 (0)	0 (0)	2 (9)	0 (0)	1 (5)	4 (18)	4 (18)	8 (36)	3 (14)	7 (1.27)
Reminders combined with social challenges (n=21)	0 (0)	0 (0)	1 (5)	0 (0)	2 (10)	3 (14)	4 (19)	6 (29)	5 (24)	7 (1.29)
Forum for general information and smartphone messages (eg, SMS text messages) for specific tailored information (n=21)	1 (5)	1 (5)	0 (0)	5 (24)	2 (10)	6 (29)	2 (10)	3 (14)	1 (5)	5 (1.67)
Forum for general information and app for specific tailored information (n=21)	1 (5)	1 (5)	1 (5)	1 (5)	4 (19)	6 (29)	1 (5)	4 (19)	2 (10)	5.5 (1.67)
Forum for general information and email for specific tailored information (n=21)	2 (10)	1 (5)	2 (10)	2 (10)	3 (14)	5 (24)	4 (19)	2 (10)	0 (0)	5 (1.71)
Website for general information and smartphone messages (eg, SMS text messages) for specific tailored information (n=23)	0 (0)	1 (4)	3 (13)	4 (17)	3 (13)	5 (22)	4 (17)	3 (13)	0 (0)	4.5 (1.65)
Website for general information and email for specific tailored information (n=23)	0 (0)	2 (9)	2 (9)	4 (17)	5 (22)	5 (22)	3 (13)	1 (4)	1 (4)	4.5 (1.52)
Website for general information and app for specific tailored information (n=23)	0 (0)	1 (4)	2 (9)	1 (4)	6 (26)	6 (26)	3 (13)	3 (13)	1 (4)	5 (1.43)

[a]MAD-M: mean absolute deviation from the median.

Round 3

After the description of the priorities of the expert panel, the BIT elements that were considered most suitable for implementing each of the work strategies are shown in Multimedia Appendix 5, including the 18 highly feasible work strategies and 16 rather than 17 highly useful BIT elements because social challenge features were considered within the gamification features in round 3.

The use of wrist-based activity trackers was identified as useful by 63% (12/19) of respondents for self-monitoring sedentary and activity behaviors. Using a computer software was considered useful to schedule 5- to 10-minute breaks between meetings (9/19, 47%), whereas a desktop application interface was considered useful to support strategies such as short breaks every 60 minutes of sitting time (6/19, 32%). Videos were described to be useful when delivering activity demonstrations to perform during the sitting breaks (9/19, 47%). Other BIT elements identified were gamification features, such as social challenges, which are used for providing awards and rewards and offering incentives (8/19, 42%). However, setting tailored goals (8/19, 42%); creating an action plan (6/19, 32%); standing while reading, answering calls, or conducting videoconferences (6/19, 32%); providing feedback on activity progress (4/19, 21%); and motivational messages (4/19, 21%) were mainly recognized to be best delivered through a smartphone app. Providing information and support was considered to be best delivered through a combination of media or a smartphone app (5/19, 26% each). Combination of media was also chosen by most of the respondents as the most useful mode for providing information to increase awareness and knowledge (5/19, 26%) and for promoting incidental movement (3/19, 16%).

Environmental strategies such as height-adjustable desks (11/19, 58%), relocation of home office supplies (5/19, 26%), standing desks (9/19, 47%), and workstation accessories (9/19, 47%) did not correspond with any BIT elements according to most of the respondents. For point-of-choice or point-of-decision prompts, there were no corresponding BIT elements to support their implementation according to most of the respondents for this strategy (4/19, 21% of the sample), although 16% (3/19) considered that this strategy could be delivered via computer software.

Focus Groups

Overview

A total of 2 indicative themes were generated from the analysis of the focus groups regarding the experts' perceptions in relation to the Delphi results on the potential feasibility and usefulness of strategies and digital elements to reduce SB in a home office context. The first theme referred to the integration of the digital strategies for reducing SB when working from home, which included 3 subthemes: social influences, physical environment, and individual factors. The second theme was related to the adaptability of the strategies to reduce SB in the office and home context.

Theme 1: Integration of the Digital Strategies for Reducing SB When Working From Home

Overview

This theme involves the importance of combining multiple levels of strategies to facilitate and maximize the strategies' implementation and their effectiveness in reducing SB when working from home. According to the experts, while organizational-level factors such as organizational support or policy changes may not be immediately feasible on their own, they are essential for achieving sustainable change. The same was considered true for information provision—it alone is not enough to motivate behavior change, although it is an important factor at the individual level to increase individuals' knowledge. In addition, environmental changes are required to promote less sitting and more movement during working hours. Therefore, experts highlighted that the following implementation levels need to work in tandem to achieve sustainable behavior change.

Social Support

Experts explained the need for the company to support the culture of active and healthy workplaces to motivate the employees' behavior change. Social support may be achieved by aligning the company policies and culture with the support of the manager (eg, motivational messages) to be able to increase the integration of digital strategies for reducing SB when working from home:

> *[I]t is based on what the management or the philosophy of the company or the company policy wants to refer to their workers. If those at the top are not proactive in...the chain, it will decrease.* [E13; focus group 4; paragraph 43]

While the target behavior is contextualized within the domain of working from home, participants identified interaction with colleagues as a relevant factor to motivate employees toward behavior change. For example, social modeling may motivate home office workers to reduce their SB by sharing activities around the sitting time breaks. However, social interaction can be perceived negatively if a competition component is present in the strategies. This may be due to competing against people whom they do not know or due to the bottom position of some workers during the challenges. Therefore, cooperative challenges may be the solution, in which every worker contributes to the challenge without a competition element:

> *[W]hen people find themselves in an environment where they've got to adapt behaviourally, seeing what can be done and observing how others might do it, can be quite powerful...we are social animals, monkey see, monkey do and that social modelling element to me jumps out as really quite important and not to be lost.* [E01; focus group 1; paragraph 44]

Physical Environment

The home office context was seen by some participants as a limiting factor in engaging in sufficient PA during work hours due to the reduced space compared to the office context. For example, having (normally) no more than 1 floor at home, implying the use of stairs, or the lack of large spaces for walking are barriers to be active while working at home. Although applying physical changes such as standing desks may be a solution to break up sitting time at work, it appears to be a less feasible strategy in the home office context due to the expensive cost for the companies:

> *I'm a big advocate of standing desks, standing workstations in terms of being effective but I appreciate that in this situation it's probably not feasible for them to be incorporated in the home-office.* [E10; focus group 3; paragraph 7]

Nevertheless, experts highlighted the importance of combining these environmental changes with prompts or reminders for the success of the behavior change integration when working from home. Regarding point-of-decision prompts, experts highlighted that mobile phone notifications were more feasible for reducing SB than desktop notifications because they cause less disturbance:

> *[W]hen we moved into our new offices a few years ago, we all got standing desks and just observing...I don't see anybody using. So, if there's something for me in there about prompting/reminders, do you know?... So, I think the prompting is really important.* [E011; focus group 3; paragraph 11]

Individual Factors

Experts mentioned the difference in literacy between *PA* and *sitting time*, believing that workers may understand the term *PA* more easily than the term *sitting time*. This may be due to the lack of knowledge on the importance of reducing sitting time behaviors, which could be perceived as a barrier when only reporting sitting time feedback. Therefore, educational material or information, one of the highlighted strategies to be used together with social support and physical environment strategies to achieve SB reduction when working from home, delivered through a combination of media (ie, smartphone, desktop, and

websites), was considered the most adaptative medium to increase the employees' knowledge. Experts emphasized that the information delivered must be simple and short, especially with the use of videos:

> *I think the combination of media is probably the way forward because I guess different organisations will also have different ways of rolling out information. I know a lot of companies nowadays that they all have their own company app, and they do a lot of communication via their app that all their employees have access to.* [E10; focus group 3; paragraph 42]

Furthermore, feedback on the respondents' own behavior was identified as a key strategy to motivate individuals and achieve behavior change. The best frequency at which to deliver feedback could be in real time as this adapts to each individual's schedule. However, a daily summary feedback approach was also considered helpful in combination with real-time feedback:

> *For the summary—daily maybe? But I probably go for the combination of the two, so you probably have some sort of summary data to share with them but still, it's nice to see at the moment what you did.* [E05; focus group 2; paragraph 67]

Theme 2: Adaptation to the Home Office Environment

This theme refers to the flexibility of the strategies to fit the diverse needs of hybrid work. All the experts mentioned the capacity of the strategies to be adapted to all environments without the need to be changed to fit in the home or office context:

> *I think you want to have the same or approach or whatever you use for say, remote working, I would also use for hybrid working because I guess a lot of strategies are going to be the same...So, I think you could maybe just replicate or duplicate what you're doing here for the hybrid workers.* [E10; focus group 3; paragraph 64]

However, some strategies need to be adapted to each individual depending on the type of work they perform. For example, walking while on a call may be a good strategy for a call center, but it might not work in another type of job. Therefore, it is valuable to provide different strategies to allow everyone to choose their preferred one to change their behavior:

> *[D]ifferent populations need a different approach, right? So very much dependent on the type of work that people are doing, you're going to have to approach them differently. So, the task set is going to help determine what options you actually have to still do your work but would be doing it sitting down all the time.* [E05; focus group 2; paragraph 29]

Therefore, experts highlighted the importance of providing a list of strategies that can be adapted to employees' needs for reducing sitting and moving more during their working hours:

> *So instead, you look towards providing a menu of strategies that you know are effective or have at least been shown to be feasible and acceptable and not shown to have any harm and you know, have benefits maybe in terms of cultural elements of it and that a lot of those would be applicable for the home environment but would also be applicable obviously for the workplace environment.* [E02; focus group 1; paragraph 65]

Summary

In general, no dissonance was found among the responses of the different experts. Strategies such as feedback, social modeling, breaks every hour, prompts via mobile phone every hour, educational material, and organizational support (eg, motivational messages) were the most discussed and considered the most feasible. Other strategies such as physical changes (ie, height-adjustable desks) were considered unfeasible despite being ranked higher in the Delphi rounds. Finally, although certain strategies such as the creation of an action plan, standing while reading and answering calls, blocking 5 to 10 minutes between meetings, or setting tailored goals for reducing sitting time obtained high scores in the Delphi rounds, these topics were discussed less among the experts during the focus groups.

Discussion

Principal Findings

This study achieved international expert consensus on 18 potentially feasible work strategies to reduce home office SB delivered through 16 potentially useful digital elements. These results seem to be transferable to both office and home settings based on experts' responses. Moreover, the experts highlighted the importance of multiple factors (ie, social, individual, and environmental) working in tandem to facilitate and maximize the strategies' implementation, as well as to achieve behavior change.

Comparison With Prior Work

Wrist-based activity trackers were identified by most of the experts (12/19, 63%) as useful for self-monitoring activity and SB. Previous research has highlighted the feasibility and usability of these devices for self-monitoring SB as well as workers' acceptability of these approaches in the office context [43,44]. In addition, these strategies have demonstrated effectiveness in reducing SB in such settings [44]. Some studies have highlighted limitations in recognition of postures and step detection when using wrist-based devices compared to leg-based devices [45-48]. However, few experts considered leg-based activity trackers for self-monitoring activity and SB due to the discomfort they may cause. This suggests that future interventions incorporating activity trackers with the aim of reducing SB should first validate the devices for their subsequent implementation and effectiveness evaluation in home office settings.

In the literature, computer visual feedback on activity patterns is considered to be an enabling factor for increasing motivation and awareness among office workers [49]. However, experts preferred smartphone apps as they are able to provide real-time feedback on activity progress. In addition, feedback may be an indication of progression toward a defined goal [43]. In this study, experts stated that smartphone apps could be a useful tool for setting tailored goals. Evidence has demonstrated the

effectiveness of smartphone apps in reducing SB and promoting PA in office settings [16]. Hence, smartphone apps may be a potential mode of delivery for setting tailored goals and providing feedback on the move, which should be evaluated in future home office environment interventions.

Previous research has demonstrated that short activity breaks (ie, walking) every 30 minutes result in reductions in office SB time [50] but also in reductions in fatigue and improvements in energy and mood [51]. However, in this study, experts considered breaks every 60 minutes more feasible. According to recent literature, a possible explanation may be related to the fact that workload and time pressure were identified by employees as the most common barriers for reducing and breaking up SB [52], making it difficult to stop more frequently. Moreover, replacing SB with active working tasks (eg, incidental moving while reading and answering calls) may overcome these barriers [49]. Working remotely increases opportunities to perform active tasks or even other light-intensity PA that may increase the potential benefits of reducing SB and its harmful health effects among office workers [9,10,53]. Experts reflected on the importance of providing clear instructions and modeling through video demonstrations as a key digital element to enhance active breaks among employees when working from home. This may have implications for future studies evaluating the most effective active break duration through video demonstrations and the implications of these breaks regarding SB time and its health effects in home office contexts.

On the basis of experts' responses, there is a literacy gap between the terms *PA* and *SB*, with activity trackers largely focused on PA instead of SB [52]. Therefore, educational material may be a key aspect to increase knowledge for behavior change and reduce the lack of understanding of SB and its subsequent health implications, which was highlighted by experts. Educational material is an essential and effective strategy for workers to increase awareness of SB's health consequences [52,54]. In our findings, a combination of media was preferred to deliver information to increase awareness and knowledge. Therefore, it is recommended that future studies include educational material to increase knowledge and awareness among home office workers.

Despite the relevance of knowledge and awareness to achieve behavior change, social and organizational support is a factor noted by workers in the current literature for facilitating and encouraging movement in the workplace [55-57], as well as by managers in the home office setting [26]. Home office work is more related to negative emotions than office work due to the isolation and loneliness; therefore, social support is even more important in this context [58]. Gamification features, including social challenges, were identified as a useful digital element to deliver strategies such as awards, rewards, or incentives to achieve goals or recommendations for reducing SB among home office workers. This strategy could provide social support and motivation. Motivational messages from managers are also an organizational support strategy, which experts suggest could be delivered through smartphone apps. Thus, future studies should evaluate the acceptability of these social and organizational strategies among both managers and employees, as well as their effectiveness in the home office context.

Environmental work strategies such as standing desks, height-adjustable desks, or workstation accessories were mostly not matched with any digital element. Moreover, experts recognized work strategies such as height-adjustable desks as unfeasible in the home office environment, primarily due to the associated costs. Recent evidence has demonstrated the high cost of such desks (eg, standing desks or height-adjustable desks) compared with nondesk interventions [59]. Workstation accessories (eg, desk platforms) or relocation of office supplies are less expensive strategies [60] that were deemed more feasible and usable by experts and may be implemented in home office settings. Therefore, home office interventions may include environmental work strategies in a similar way to office interventions but with a lower associated cost.

On the basis of the findings presented in this paper, physical environmental strategies need to work in tandem with social support and individual strategies to achieve behavior change. A recent systematic review showed that theory-based interventions, a relevant step in multicomponent intervention development [25], that include social, physical, and individual elements are more effective than non–theory-based interventions [21]. Furthermore, stakeholder engagement (eg, consultation or coproduction among researchers, experts on the topic, and the target population) should be encouraged throughout the development process [25]. Recent studies have highlighted the need for a top-down approach whereby managers should provide support to employers [26]; hence, the implementation of strategies should be an organizational initiative. Although evaluating effectiveness on outcomes has traditionally been the primary focus of interventions in this discipline, in particular to encourage their adoption in the real world, the feasibility, acceptability, cost-effectiveness, and scalability of interventions is increasingly important to ensure the translation of findings from research to practice [25,61]. This study provides new insights for future research on the design of home office interventions aimed at reducing SB based on the learned experiences of occupational health experts. However, we were not able to cover the entire audience's perspective, so new qualitative studies exploring the perspectives of different stakeholders (eg, end users) may further inform about the feasibility and usability of digital workplace strategies.

In addition, experts considered that no adaptations would be needed to transfer these home office strategies to hybrid work, the current modus operandi for many desk-based workers. However, the experts who participated in this study were mainly from high-income countries; therefore, such perceptions may vary among people from other regions. Hybrid work presents opportunities for promoting PA by providing flexibility to integrate occupational, leisure, lifestyle, and incidental activities [9] but also for improving health and well-being [62]. Hence, the results of this study may be transferable to hybrid work in high-income countries, whereas hybrid work should receive greater consideration in future interventions given cultural variability.

Strengths and Limitations

The main strength of this study is the novelty of the home office context to deliver strategies for reducing SB through digital

elements. The fact that focus groups were conducted after a Delphi study enabled us to elicit additional views and opinions on the Delphi results from the experts' perspectives, consolidating the consensus. Nevertheless, we did not share the analytical categories, interpretations, or conclusions from the thematic analysis with the experts, which may affect the analysis's credibility. The entire study was conducted in a rigorous manner, with high participant response rates and representation from international experts. However, some strategies may have been overlooked due to the nonsystematic search approaches used, and most of the international experts were from high-income countries, limiting the applicability of these findings to those communities. Some other limitations were present in the study. First, despite the rigor in the methodology used to engage experts, not all responses were complete, meaning that data were sometimes missing from some questions. Second, as is the case in any Delphi survey, the data gathered were based on subjective opinion and the expertise of the participants. Third, experts acted as representatives of the stakeholders; thus, the inclusion of other stakeholders such as end users would also have been important to address other, more diverse perspectives.

Conclusions

This study presents the consensus of international experts on feasible work strategies and useful digital elements as a mode of delivery of these work strategies to support the reduction in SB in home office or hybrid contexts. Consensus was achieved on 18 work strategies and 16 BIT elements. Future interventions should implement and evaluate the effectiveness of these strategies but also their feasibility and acceptability in home office settings.

Acknowledgments

The authors thank all the participants for contributing their time and expertise, including Dianne Commissaris, Emma George, Erwin Speklé, Frank Vandaele, Genevieve Healy, Ida Damen, Ilkka Väänänen, Ingrid Hendriksen, James McKenna, Judith Jelsma, Katja Wouters, Leon Straker, Maja Pajeck, Matthew P Buman, Mette Jensen Stochkendahl, Neville Owen, Nipun Shrestha, Pania Karnaki, Pieter Coenen, Satu Mänttäri, Sebastià Mas Alòs, Stacy Clemes, Wendy J Brown, and others with their names withheld. This research was funded by the Cooperation Partnerships program of the European Commission (101050490-ERASMUS-SPORT-2021-SCP).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Delphi surveys. The document shows the completed survey of round 1 of the Delphi study on the first page. On the second page, the file presents the completed survey of round 2, and the third page shows the completed survey of round 3.
[XLSX File (Microsoft Excel File), 37 KB - humanfactors_v12i1e59405_app1.xlsx]

Multimedia Appendix 2
Focus group guide. The file shows the complete guide with the times of the focus groups.
[DOCX File , 22 KB - humanfactors_v12i1e59405_app2.docx]

Multimedia Appendix 3
Feasibility of the work strategies (results of round 1).
[DOCX File , 66 KB - humanfactors_v12i1e59405_app3.docx]

Multimedia Appendix 4
Usefulness of behavioral intervention technology elements (results of round 1).
[DOCX File , 22 KB - humanfactors_v12i1e59405_app4.docx]

Multimedia Appendix 5
Matching behavioral intervention technology elements with work strategies on which consensus was achieved (round 3 results).
[DOCX File , 26 KB - humanfactors_v12i1e59405_app5.docx]

References

1. Patel AK, Banga C, Chandrasekaran B. Effect of an education-based workplace intervention (move in office with education) on sedentary behaviour and well-being in desk-based workers: a cluster randomized controlled trial. Int J Occup Saf Ergon 2022 Sep;28(3):1655-1663. [doi: 10.1080/10803548.2021.1916221] [Medline: 33843490]

2. Puig-Ribera A, Martínez-Lemos I, Giné-Garriga M, González-Suárez Á, Bort-Roig J, Fortuño J, et al. Self-reported sitting time and physical activity: interactive associations with mental well-being and productivity in office employees. BMC Public Health 2015 Jan 31;15:72 [FREE Full text] [doi: 10.1186/s12889-015-1447-5] [Medline: 25886270]
3. Rosenkranz SK, Mailey EL, Umansky E, Rosenkranz RR, Ablah E. Workplace sedentary behavior and productivity: a cross-sectional study. Int J Environ Res Public Health 2020 Sep 08;17(18):6535 [FREE Full text] [doi: 10.3390/ijerph17186535] [Medline: 32911740]
4. Putsa B, Jalayondeja W, Mekhora K, Bhuanantanondh P, Jalayondeja C. Factors associated with reduced risk of musculoskeletal disorders among office workers: a cross-sectional study 2017 to 2020. BMC Public Health 2022 Aug 06;22(1):1503 [FREE Full text] [doi: 10.1186/s12889-022-13940-0] [Medline: 35932005]
5. Global action plan on physical activity 2018-2030: more active people for a healthier world. World Health Organization. 2019. URL: https://www.who.int/publications-detail-redirect/9789241514187 [accessed 2023-10-02]
6. Waters CN, Ling EP, Chu AH, Ng SH, Chia A, Lim YW, et al. Assessing and understanding sedentary behaviour in office-based working adults: a mixed-method approach. BMC Public Health 2016 Apr 27;16:360 [FREE Full text] [doi: 10.1186/s12889-016-3023-z] [Medline: 27117178]
7. Keown MK, Skeaff CM, Perry TL, Haszard JJ, Peddie MC. Device-measured sedentary behavior patterns in office-based university employees. J Occup Environ Med 2018 Dec;60(12):1150-1157. [doi: 10.1097/JOM.0000000000001467] [Medline: 30308622]
8. Wiseman B. The next great disruption is hybrid work—are we ready? Microsoft WorkLab. 2021. URL: https://www.brianheger.com/2021-work-trend-index-annual-report-the-next-great-disruption-is-hybrid-work-are-we-ready-microsoft/ [accessed 2024-04-29]
9. Gilson N, Coenen P, Hallman D, Holtermann A, Mathiassen SE, Straker L. Postpandemic hybrid work: opportunities and challenges for physical activity and public health. Br J Sports Med 2022 Nov;56(21):1203-1204. [doi: 10.1136/bjsports-2022-105664] [Medline: 35710111]
10. Smite D, Moe NB, Hildrum J, Huerta JG, Mendez D. Work-from-home is here to stay: call for flexibility in post-pandemic work policies. J Syst Softw 2023 Jan;195:111552 [FREE Full text] [doi: 10.1016/j.jss.2022.111552] [Medline: 36320721]
11. Loef B, van Oostrom SH, Bosma E, Lifelines Corona Research Initiative, Proper KI. The mediating role of physical activity and sedentary behavior in the association between working from home and musculoskeletal pain during the COVID-19 pandemic. Front Public Health 2022;10:1072030 [FREE Full text] [doi: 10.3389/fpubh.2022.1072030] [Medline: 36530694]
12. Tronco Hernández YA, Parente F, Faghy MA, Roscoe CM, Maratos FA. Influence of the COVID-19 lockdown on the physical and psychosocial well-being and work productivity of remote workers: cross-sectional correlational study. JMIRx Med 2021;2(4):e30708 [FREE Full text] [doi: 10.2196/30708] [Medline: 34898665]
13. Pereira MA, Mullane SL, Toledo MJL, Larouche ML, Rydell SA, Vuong B, et al. Efficacy of the 'Stand and Move at Work' multicomponent workplace intervention to reduce sedentary time and improve cardiometabolic risk: a group randomized clinical trial. Int J Behav Nutr Phys Act 2020 Oct 27;17(1):133 [FREE Full text] [doi: 10.1186/s12966-020-01033-3] [Medline: 33109190]
14. Larisch L, Blom V, Hagströmer M, Ekblom M, Ekblom Ö, Nilsson J, et al. Improving movement behavior in office workers: effects of two multi-level cluster-RCT interventions on mental health. BMC Public Health 2024 Jan 09;24(1):127 [FREE Full text] [doi: 10.1186/s12889-024-17647-2] [Medline: 38195449]
15. Huang Y, Benford S, Blake H. Digital interventions to reduce sedentary behaviors of office workers: scoping review. J Med Internet Res 2019 Feb 07;21(2):e11079 [FREE Full text] [doi: 10.2196/11079] [Medline: 30730294]
16. Buckingham SA, Williams AJ, Morrissey K, Price L, Harrison J. Mobile health interventions to promote physical activity and reduce sedentary behaviour in the workplace: a systematic review. Digit Health 2019;5:2055207619839883 [FREE Full text] [doi: 10.1177/2055207619839883] [Medline: 30944728]
17. Edwardson CL, Yates T, Biddle SJ, Davies MJ, Dunstan DW, Esliger DW, et al. Effectiveness of the Stand More AT (SMArT) work intervention: cluster randomised controlled trial. BMJ 2018 Oct 10;363:k3870 [FREE Full text] [doi: 10.1136/bmj.k3870] [Medline: 30305278]
18. Kitano N, Kai Y, Jindo T, Fujii Y, Tsunoda K, Arao T. Association of domain-specific physical activity and sedentary behavior with cardiometabolic health among office workers. Scand J Med Sci Sports 2022 Aug;32(8):1224-1235. [doi: 10.1111/sms.14165] [Medline: 35426181]
19. Olsen HM, Brown WJ, Kolbe-Alexander T, Burton NW. A brief self-directed intervention to reduce office employees' sedentary behavior in a flexible workplace. J Occup Environ Med 2018 Oct;60(10):954-959. [doi: 10.1097/JOM.0000000000001389] [Medline: 30001255]
20. Wilkerson AH, Elliott CR, McFadden NT, Abutalib N. Feasibility of using mobile standing desks to address sedentary behavior in flexible work environments: a mixed methods study. J Occup Environ Med 2023 May 01;65(5):e273-e278 [FREE Full text] [doi: 10.1097/JOM.0000000000002804] [Medline: 36701795]
21. Parés-Salomón I, Señé-Mir AM, Martín-Bozas F, Loef B, Coffey A, Dowd KP, et al. Effectiveness of workplace interventions with digital elements to reduce sedentary behaviours in office employees: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2024 Apr 19;21(1):41 [FREE Full text] [doi: 10.1186/s12966-024-01595-6] [Medline: 38641816]

22. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
23. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
24. Kwasnicka D, Dombrowski SU, White M, Sniehotta F. Theoretical explanations for maintenance of behaviour change: a systematic review of behaviour theories. Health Psychol Rev 2016 Sep;10(3):277-296 [FREE Full text] [doi: 10.1080/17437199.2016.1151372] [Medline: 26854092]
25. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
26. Coffey A, Parés-Salomón I, Reckman P. Cross-cultural needs analysis: factors for reducing sedentary behaviours among home-office workers. Click2Move Project. URL: https://click2moveproject.com/phases/phase-1/ [accessed 2024-04-08]
27. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]
28. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. PLoS One 2011;6(6):e20476 [FREE Full text] [doi: 10.1371/journal.pone.0020476] [Medline: 21694759]
29. Jünger S, Payne SA, Brine J, Radbruch L, Brearley SG. Guidance on conducting and REporting DElphi studies (CREDES) in palliative care: recommendations based on a methodological systematic review. Palliat Med 2017 Sep 13;31(8):684-706. [doi: 10.1177/0269216317690685] [Medline: 28190381]
30. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
31. Väänänen I, Mas-Alòs S, Vandaele F, Codina-Nadal A, Matas S, Aumatell E, et al. Workplace physical activity practices in real life: a scoping review of grey literature for small- and medium-sized enterprises. Eur J Public Health 2022 Aug 26;32(Suppl 1):i22-i27 [FREE Full text] [doi: 10.1093/eurpub/ckac083] [Medline: 36031820]
32. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
33. Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011 Apr;64(4):395-400. [doi: 10.1016/j.jclinepi.2010.09.012] [Medline: 21194891]
34. Gadke DL, Kratochwill TR, Gettinger M. Incorporating feasibility protocols in intervention research. J Sch Psychol 2021 Feb;84:1-18. [doi: 10.1016/j.jsp.2020.11.004] [Medline: 33581765]
35. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 05;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
36. Elamir E. Mean absolute deviation about median as a tool of explanatory data analysis. Int J Res Rev Appl Sci 2012;11:e630 [FREE Full text]
37. Tran TN, Felfernig A, Le VM. An overview of consensus models for group decision-making and group recommender systems. User Model User-adapt Interact 2024;34(3):489-547 [FREE Full text]
38. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
39. Braun V, Clarke V. Toward good practice in thematic analysis: avoiding common problems and be(com)ing a researcher. Int J Transgend Health 2023;24(1):1-6 [FREE Full text] [doi: 10.1080/26895269.2022.2129597] [Medline: 36713144]
40. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019;11(4):589-597 [FREE Full text]
41. Kakar ZU, Rasheed R, Rashid A, Akhter S. Criteria for assessing and ensuring the trustworthiness in qualitative research. Int J Bus Reflect 2023 Dec 31;4(2):150-173. [doi: 10.56249/ijbr.03.01.44]
42. Ahmed SK. The pillars of trustworthiness in qualitative research. J Med Surg Public Health 2024:2 [FREE Full text]
43. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
44. Guitar NA, MacDougall A, Connelly DM, Knight E. Fitbit activity trackers interrupt workplace sedentary behavior: a new application. Workplace Health Saf 2018 May 09;66(5):218-222. [doi: 10.1177/2165079917738264] [Medline: 29121833]
45. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]

46. Tudor-Locke C, Leonardi C, Johnson WD, Katzmarzyk PT. Time spent in physical activity and sedentary behaviors on the working day: the American time use survey. J Occup Environ Med 2011 Dec;53(12):1382-1387. [doi: 10.1097/JOM.0b013e31823c1402] [Medline: 22104979]
47. Rosenberger ME, Haskell WL, Albinali F, Mota S, Nawyn J, Intille S. Estimating activity and sedentary behavior from an accelerometer on the hip or wrist. Med Sci Sports Exerc 2013 May;45(5):964-975 [FREE Full text] [doi: 10.1249/MSS.0b013e31827f0d9c] [Medline: 23247702]
48. Fanning J, Miller M, Chen S, Davids C, Kershner K, Rejeski W. Is wrist accelerometry suitable for threshold scoring? A comparison of hip-worn and wrist-worn actigraph data in low-active older adults with obesity. J Gerontol A Biol Sci Med Sci 2022 Dec 29;77(12):2429-2434 [FREE Full text] [doi: 10.1093/gerona/glab347] [Medline: 34791237]
49. Bort-Roig J, Martin M, Puig-Ribera A, González-Suárez Á, Martínez-Lemos I, Martori JC, et al. Uptake and factors that influence the use of 'sit less, move more' occupational intervention strategies in Spanish office employees. Int J Behav Nutr Phys Act 2014 Dec 10;11:152 [FREE Full text] [doi: 10.1186/s12966-014-0152-6] [Medline: 25490857]
50. Mailey EL, Rosenkranz SK, Casey K, Swank A. Comparing the effects of two different break strategies on occupational sedentary behavior in a real world setting: a randomized trial. Prev Med Rep 2016 Dec;4:423-428 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.010] [Medline: 27583200]
51. Mailey EL, Rosenkranz SK, Ablah E, Swank A, Casey K. Effects of an intervention to reduce sitting at work on arousal, fatigue, and mood among sedentary female employees: a parallel-group randomized trial. J Occup Environ Med 2017 Dec;59(12):1166-1171. [doi: 10.1097/JOM.0000000000001131] [Medline: 28816735]
52. Edwardson CL, Maylor BD, Biddle SJ, Clarke-Cornwell AM, Clemes SA, Davies MJ, et al. Participant and workplace champion experiences of an intervention designed to reduce sitting time in desk-based workers: SMART work and life. Int J Behav Nutr Phys Act 2023 Nov 30;20(1):142 [FREE Full text] [doi: 10.1186/s12966-023-01539-6] [Medline: 38037043]
53. Del Pozo-Cruz J, García-Hermoso A, Alfonso-Rosa RM, Alvarez-Barbosa F, Owen N, Chastin S, et al. Replacing sedentary time: meta-analysis of objective-assessment studies. Am J Prev Med 2018 Sep;55(3):395-402. [doi: 10.1016/j.amepre.2018.04.042] [Medline: 30122216]
54. Edwardson CL, Abell L, Clarke-Cornwell A, Dunstan DW, Gray LJ, Healy GN, et al. Implementation and engagement of the SMART Work and Life sitting reduction intervention: an exploratory analysis on intervention effectiveness. Int J Behav Nutr Phys Act 2023 Dec 19;20(1):148. [doi: 10.1186/s12966-023-01548-5] [Medline: 38115044]
55. Kim J, Mizushima R, Nishida K, Morimoto M, Nakata Y. Proposal of a comprehensive and multi-component approach to promote physical activity among Japanese office workers: a qualitative focus group interview study. Int J Environ Res Public Health 2022 Feb 15;19(4):2172 [FREE Full text] [doi: 10.3390/ijerph19042172] [Medline: 35206358]
56. Taylor WC, Suminski RR, Das BM, Paxton RJ, Craig DW. Organizational culture and implications for workplace interventions to reduce sitting time among office-based workers: a systematic review. Front Public Health 2018;6:263 [FREE Full text] [doi: 10.3389/fpubh.2018.00263] [Medline: 30320051]
57. Mackenzie K, Such E, Norman P, Goyder E. Sitting less at work: a qualitative study of barriers and enablers in organisations of different size and sector. BMC Public Health 2019 Jul 04;19(1):884 [FREE Full text] [doi: 10.1186/s12889-019-7148-8] [Medline: 31272484]
58. Beckel JL, Fisher GG. Telework and worker health and well-being: a review and recommendations for research and practice. Int J Environ Res Public Health 2022 Mar 24;19(7):3879 [FREE Full text] [doi: 10.3390/ijerph19073879] [Medline: 35409563]
59. Cox E, Walker S, Edwardson CL, Biddle SJ, Clarke-Cornwell AM, Clemes SA, et al. The cost-effectiveness of the SMART work and life intervention for reducing sitting time. Int J Environ Res Public Health 2022 Nov 11;19(22):14861 [FREE Full text] [doi: 10.3390/ijerph192214861] [Medline: 36429578]
60. Munir F, Miller P, Biddle SJ, Davies MJ, Dunstan DW, Esliger DW, et al. A cost and cost-benefit analysis of the stand more at work (SMArT work) intervention. Int J Environ Res Public Health 2020 Feb 13;17(4):1214 [FREE Full text] [doi: 10.3390/ijerph17041214] [Medline: 32070034]
61. Skivington K, Matthews L, Simpson SA, Craig P, Baird J, Blazeby JM, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. BMJ 2021 Sep 30;374:n2061 [FREE Full text] [doi: 10.1136/bmj.n2061] [Medline: 34593508]
62. Olsen HM, Brown WJ, Kolbe-Alexander T, Burton NW. Flexible work: the impact of a new policy on employees' sedentary behavior and physical activity. J Occup Environ Med 2018 Jan;60(1):23-28. [doi: 10.1097/JOM.0000000000001190] [Medline: 29023342]

Abbreviations

BIT: behavioral intervention technology
COREQ: Consolidated Criteria for Reporting Qualitative Research
MAD-M: mean absolute deviation from the median
PA: physical activity
SB: sedentary behavior

Edited by A Kushniruk; submitted 24.04.24; peer-reviewed by A Wagg, SA Alothman; comments to author 09.01.25; revised version received 12.02.25; accepted 04.03.25; published 08.04.25.

Please cite as:
Parés-Salomón I, Vaqué-Crusellas C, Coffey A, Loef B, Proper KI, Señé-Mir AM, Puig-Ribera A, Dowd KP, Bort-Roig J
Development of Digital Strategies for Reducing Sedentary Behavior in a Hybrid Office Environment: Modified Delphi Study
JMIR Hum Factors 2025;12:e59405
URL: https://humanfactors.jmir.org/2025/1/e59405
doi:10.2196/59405
PMID:40198908

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org