JMIR Human Factors

Impact Factor (2022): 2.7
Volume 8 (2021), Issue 1 ISSN 2292-9495 Editor in Chief: Andre Kushniruk, BA, MSc, PhD, FACMI

Contents

Original Papers

Reviews

Original Paper

A Web-Based Application for Complex Health Care Populations: User-Centered Design Approach

Francesca Ferrucci[1,2,3], PhD; Manuele Jorio[1], PsyD; Stefano Marci[4], MD; Antonia Bezenchek[1,3], MScIT; Giulia Diella[5], NP; Cinzia Nulli[4], NP; Ferdinando Miranda[5], NP; Guido Castelli-Gattinara[5], MD

[1]Informapro Srl, Rome, Italy

[2]Department of Human Sciences, Communication and Tourism, University of Tuscia, Viterbo, Italy

[3]EuResist Network European Economic Interest Grouping, Rome, Italy

[4]Unità Operativa Complessa Materno-Infantile - Azienda Sanitaria Locale Rieti, Consultorio Pediatrico, Rieti, Italy

[5]Academic Department of Pediatrics, Division of Immune and Infectious Diseases - Istituto di Ricovero e Cura a Carattere Scientifico Ospedale Pediatrico Bambino Gesù, Rome, Italy

Corresponding Author:
Francesca Ferrucci, PhD
Informapro Srl
via Guido Guinizelli, 98/100
Rome, 00152
Italy
Phone: 39 06 5502218
Email: fferrucci@unitus.it

Abstract

Background: Although eHealth technology makes it possible to improve the management of complex health care systems and follow up on chronic patients, it is not without challenges, thus requiring the development of efficient programs and graphic user interface (GUI) features. Similar information technology tools are crucial, as health care populations are going to have to endure social distancing measures in the forthcoming months and years.

Objective: This study aims to provide adequate and personalized support to complex health care populations by developing a specific web-based mobile app. The app is designed around the patient and adapted to specific groups, for example, people with complex or rare diseases, autism, or disabilities (especially among children) as well as Alzheimer or senile dementia. The app's core features include the collection, labeling, analysis, and sorting of clinical data. Furthermore, it authorizes a network of people around the patient to securely access the data contained in his or her electronic health record.

Methods: The application was designed according to the paradigms of patient-centered care and user-centered design (UCD). It considers the patient as the main empowered and motivating factor in the management of his or her well-being. Implementation was informed through a family needs and technology perception assessment. We used 3 interdisciplinary focus groups and 2 assessment surveys to study the contexts of app use, subpopulation management, and preferred functions. Finally, we developed an observational study involving 116 enrolled patients and 253 system users, followed by 2 feedback surveys to evaluate the performance and impact of the app.

Results: In the validated general GUI, we developed 10 user profiles with different privacy settings. We tested 81 functions and studied a modular structure based on disease or medical area. This allowed us to identify replicable methods to be applied to module design. The observational study not only showed good family and community engagement but also revealed some limitations that need to be addressed. In total, 42 of 51 (82%) patients described themselves as *satisfied* or *very satisfied*. Health care providers reported facilitated communication with colleagues and the need to support data quality.

Conclusions: The experimented solution addressed some of the health system challenges mentioned by the World Health Organization: usability appears to be significantly improved when the GUI is designed according to patients' UCD mental models and when new media and medical literacy are promoted. This makes it possible to maximize the impact of eHealth products, thereby overcoming some crucial gaps reported in the literature. Two main features seemed to have potential benefit compared with other eHealth products: the modeling, within the app, of both the formal and informal health care support networks and the modular structure allowing for comorbidity management, both of which require further implementation.

(JMIR Hum Factors 2021;8(1):e18587) doi:10.2196/18587

KEYWORDS

patient; community participation; eHealth; patient-centered care; user-centered design, comorbidity

Introduction

Background

The improvement in health services and the quality of health treatment and social care has led to a significant increase in survival (and quality of life) among adults and children with chronic complex diseases and high health care needs [1].

According to the World Health Organization (WHO), over a billion people have some form of disability, whereas 110 to 190 million adults have significant difficulty functioning. An estimated 39% of the Italian population is affected by some chronic disease, with increasing disability rates. Currently, more than 3 million people in Italy are disabled. These patients are characterized by multiple morbidities, requiring the use of a range of services and a technology-enhanced care model [1-4].

eHealth may help such patients manage multiple clinical encounters and large amounts of clinical information generated from various sources. Indeed, patients report a highly frequent use of information and communications technology (ICT) to search for health information, communicate with health care providers (HCPs), track medical information and medications, and assist in decision making regarding treatment [5]. Notably, patients attempt to use ICT tools for self-management, as they expect to benefit from eHealth and enhance control over their own disease [6].

Extant research suggests that eHealth tools supporting patient-HCP interaction, patient self-management, and HCP-HCP interactions (through electronic health record integration) are of great benefit to patients [7,8]. These benefits may increase further, as the COVID-19 crisis has triggered additional demand for remote care models and systems. Previous studies have pointed out a number of critical issues concerning complex health care populations, since these include different subpopulations that pose specific medical and organizational challenges for the design of public service provision. These issues include the accurate assessment of the levels of services and needs, implementation of services and resources tailored to specific needs, coordination and integration of family-centered care planning, promotion of health systems based on patient or family self-management, and the redefinition of models of multidisciplinary team care [5,9,10].

According to the 2012-2020 eHealth Action Plan, in 2011, the Italian Public Administration promoted a high-communication health care project and a citizen's Electronic Health Dossier (*Fascicolo Sanitario Elettronico*) [8,11], but the project encountered difficulties in getting under way and proved difficult to implement. The few ongoing initiatives have not received positive feedback from users due to usability problems and the low digital literacy of both HCPs and families [12].

Objectives

In this context, the ABILITA2 Project (Italian: *Sviluppo di un Applicativo per terminali moBILI dedicato a popolazioni ad alTA complessità Assistenziale*; English: *Development of a web-based Mobile Application for complex healthcare populations*) takes advantage of ICT and its eHealth applications, exploiting the patient-centered care approach. When addressing the abovementioned issues, it adapts the service to different subpopulations, providing models that can be replicated in the future [13].

To meet the requirement of interdisciplinarity, the ABILITA2 consortium includes a partnership between ICT companies (Informapro Srl, Logica Informatica Srl, and Mediamed Interactive Srl) and medical and research centers (*Ospedale Pediatrico Bambino Gesù* - Rome and *Consultorio Pediatrico* ASL Rieti) as well as patient associations related to the medical areas of Alzheimer disease, autism, artificial nutrition, and rare pediatric diseases.

The project's general objective was to provide adequate and personalized support to complex health care populations by developing a specific web-based app, *Abilita*, designed around the patient and customizable for specific groups, notably people with complex or rare diseases (eg, genetic syndromes, patients requiring parenteral nutrition), autism or disabilities (especially among children), and Alzheimer or senile dementia. The core features of the app allow for the collection, labeling, analysis, and sorting of clinical data. Furthermore, it authorizes a network of people around the patient to securely access the data contained in his or her electronic health record.

The study's specific objectives are as follows:

- Assess levels of service and patient needs, testing assessment procedures and tools, especially for pediatric and older adult groups who are less considered in the eHealth market.
- Promote patient self-management and co-responsibility as the basis for a suitable and user-friendly web application. The emphasis is on patient empowerment (understanding of his or her role, acquisition of sufficient knowledge to be able to engage with HCPs, patient skills, and the availability of a facilitating environment [14,15]).
- Enhance and innovate the coordination between professionals and caregivers, specifically exploring the potential of a collaborative network operating on the patient's behalf, which is built by the patient based on his or her individual needs and institutional contacts.
- Make the most of a *proximity support network*, which includes informal relationships with relatives, friends, and key figures in the territory, which is a crucial health care management factor [16,17].
- Encourage families or communities to play an active role and, at the same time, ensure quality of data, care, and

assistance by using GUI modeling of proper actions per profile according to the level of skill and motivation.
- Assess the app's performance and impact.

Methods

Assessment and Design Process

The project adopted a user-centered design (UCD) approach in graphic user interfaces (GUIs) and considered users' point of view and needs as central. The difference from other methods is that UCD meets the needs and desires of users rather than forcing them to change their behavior to meet the product settings [18]. Since the designers considered the user to be the patient (or parent/caregiver), an interdisciplinary analysis was needed to assess needs and then model actions, logic paths, questions, and answers within the interface. To do so, clinical and medical competence needs to be flanked by skills in computer sciences and database management, communication or new media sciences, psychology, and sociology [13]. The study used a number of focus groups based on a general inductive approach. The results of these focus groups were then further investigated through anonymous questionnaires [19]. The focus groups met monthly with 90- to 120-min sessions to analyze the different issues raised by the study.

Focus group A assessed patients' needs and scenarios of use. It included patients (n=4), health care workers (n=2), psychologists (n=1), researchers in communication sciences (n=1), and software developers (n=1). All participants were part of the project network and discussed the experience of patients and caregivers with ICT products and possible scenarios using the Abilita app. Finally, a web-based questionnaire (Q1) was developed for the purpose of studying the main features, habits, needs, and digital and medical literacy of patients and families. Q1 was sent to a selected sample of patient associations (presidents and expert members in steering groups): *Alzheimer Uniti Roma ONLUS, Associazione Nazionale Genitori Soggetti Autistici (ANGSA) Lazio Onlus, Associazione italiana sulla nutrizione Artificiale Domiciliare "Un filo per la Vita," Associazione Prader Willi Lazio, Associazione Italiana delezione cromosoma 22 Onlus*. The 20 anonymous responses were collected in June 2018; and the statistics of multiple-choice items and summaries of open-answer items were contained in a project report in September 2018 [20,21].

Focus group B, consisting of HCPs (n=4), psychologists (n=1), privacy officers (n=1), and software developers (n=2), was devoted to the general GUI design. The outcomes of the assessment of patient needs were translated into design challenges. The discussion raised a number of research questions, including the problem of low HCP motivation or time and the need to consider the patient as the main subject motivated to use the app. It is also necessary to task the patient or caregiver with data entry and updating health records and adding user profiles to the app (to model both institutional and informal patient support networks). Additional issues concerned the powers of individual user profiles (reading or writing of sections of the data set), the need to ensure health data quality, even when not directly entered by HCPs, and to predict real-world data entered by the patient and his or her proximity network. We used paper prototyping throughout the process that led to the user requirements document delivered in November 2018 for all identified user profiles (patient, parent or tutor, caregiver, family member, doctor, nurse, structure manager, social operator, temporary, and emergency).

In designing the health record, we tried to identify possible user behaviors, which led to additional questions: what does a particular population require and how can the interface structure be customized for specific pathologies to meet patient needs and coordination requirements? Data and pages are not equally relevant for all subpopulations, and preferred content, information, and functionalities differ across groups. In this respect, the general GUI of *Abilita* could be made more powerful by customizing content and database structure, with a view to create GUIs for more specific medical areas (the Abilita *modules*).

Focus group C was set up to assess this potential. It included presidents and steering group members from patient associations (n=4), psychologists (n=1), communication sciences researchers (n=1), and software developers (n=1). The discussion addressed the specific needs of the subpopulations involved in the study, after which we administered a mandatory questionnaire (Q2) to test the usefulness and effectiveness of feasible implementations. Q2 was sent out through email to a selected sample of national and regional patient associations; the 15 anonymous responses were then collected into a database highlighting the main aspects or attention points for GUI customization and the preferred functions that could be identified.

Observational Study, Feedback, and Validation

After the development of the prototype, we performed an observational study to evaluate its application in terms of its functionality, versatility, responsiveness to patients or families' needs, user-friendliness, and rate of acceptance. We designed the study in line with international Good Clinical Practice criteria and obtained approval from the ethics committees of the medical centers involved (document protocols 1589_OPBG_2018 and 2474/CE Lazio1).

A total of 116 of the 130 (89.2%) patients invited to participate in the study were included, as they (or their families) possessed the required computer skills. They were recruited in the Rome area and in the Province of Rieti, a setting marked by a variety of health needs and increased geographic isolation due to the 2016 earthquake. During the 6-month study period (January-June 2019), the patients authorized additional user profiles to access their data, namely 32 HCPs, 97 parents, 5 family members, and 3 caregivers, for a total of 253 app users.

We then analyzed individual user accesses to explore the actual use of the app. Frontal, telephone, and web-based tutoring sessions helped the patient participants (or their parents if the patient was aged under 16 years) to complete the registration and browse the app upon uploading their personal data. In June 2019, we developed a voluntary web application feedback questionnaire for patients (Q3) with indicators for evaluating usefulness or satisfaction, privacy, and security impact. We identified usability and effectiveness, while task managers tested

the app's compliance with general recommendations and technical functionality. A link to the questionnaire was sent by email (we avoided multiple responses by limiting survey access to a single instance), and we received 51 anonymous responses in July 2019; the statistics on multiple-choice items and summaries of open-answer items were reported in a project report in September 2019.

In July 2019, we conducted 23 semistructured individual interviews with 10 doctors and 13 nurses to explore the app's usefulness in the follow-up of chronic patients, its usability, and other features of the HCP interface (questionnaire Q4).

Table 1 summarizes the different data collection stages of the research.

Table 1. Data collection processes.

Data collection process	Description	Access and recruitment criteria	Collected data and period	Output
Focus group A	8 participants (4 members of the patients' associations or caregivers, 2 HCPs[a], 1 software programmer, and 1 psychologist); 1 facilitator (researcher in communication sciences)	Members of the project network, experienced in the management of 5 medical areas (autism spectrum disorders, 22q11.2 deletion syndrome, Alzheimer disease, Prader-Willi syndrome, chronic intestinal failure)	Eight 2-hour meetings in the period, April-May 2018	Definition of main aspects and attention points to be tested on a larger sample of respondents through the questionnaire Q1; definition of scenarios of use
Questionnaire Q1	62 items mostly in a multiple-choice format and with partial adaptative questioning	A web questionnaire mandatory for a restricted sample of national and regional patient association members (closed mandatory survey [21])	20 anonymous responses collected in May 2018	Project report
Focus group B	8 participants (2 software programmers, 2 doctors, 2 nurses, 1 psychologist, and 1 privacy officer); 1 facilitator (researcher in communication sciences)	Members of the project network, experienced in eHealth and GUI[b] design processes	Fifteen 2-hour meetings in the period, June-November 2018	User requirement document for all the identified user profiles
Focus group C	6 participants (4 members of the patients' associations, 1 software programmer, and 1 psychologist); 1 facilitator (researcher in communication sciences)	Members of the project network, experienced in the management of 5 medical areas (autism spectrum disorders, 22q11.2 deletion syndrome, Alzheimer disease, Prader-Willi syndrome, chronic intestinal failure)	Five 2-hour meetings in the period, December 2018-January 2019	Definition of main aspects and attention points to be tested on a larger sample of respondents through the questionnaire Q2
Questionnaire Q2	7 items mostly in an open-answer format	Text file sent by email to a selected sample of national and regional patient association members (closed mandatory survey [21])	15 anonymous responses collected in January 2019	Database with main aspects and attention points for customization of the GUI
Observational study	Use of the *Abilita* app in real-world settings by patients, families, HCPs, and communities	We invited 130 patients of the project medical centers to participate (Provinces of Rome and Rieti); 116 accepted the invitation and were recruited	253 system users in the period January-June 2019 (116 patients, 32 HCPs, 97 parents, 5 other family members, and 3 caregivers)	Report on statistics of use in real-world settings exported by the system administrators
Questionnaire Q3	36 items mostly in a multiple-choice format (16 defined by a Likert scale score) and with partial adaptative questioning	A web questionnaire; we invited the 116 patients involved in the observational study and obtained 51 responses (closed voluntary survey [21])	51 responses collected in July 2019	Project report
Questionnaire Q4	17 items (16 defined by a Likert scale score and 1 open-answer item)	Face-to-face interviews; we invited the 32 HCPs involved in the observational study; 23 accepted	23 responses collected in July 2019	Project report

[a]HCP: health care provider.

[b]GUI: graphic user interface.

Results

Assessment and Design Process

Q1 clarified the overall context of the study. The age at first diagnosis for complex health care diseases ranged from 0 to 5 years for the majority of cases and from 65 to 80 years in the remaining cases. All patients were not autonomous and had at least one caregiver. Their digital skills were at a basic or medium level, with limited experience with the use of IT tools to communicate with social and (private or public) health care services. Patients or caregivers displayed significant awareness of their medical areas. They were able to name the diagnosis in technical terms, describe the main elements of the disorder or disease (causes, severity, symptomatology, correlations with other disorders, and risk factors), mention the pharmacological therapies with precision, describe recommended daily treatments and activities (diets, sport), and recognize changes in symptoms (especially aspects to be monitored and reported to health care personnel). The most frequently used documents were treatment plans, reports of visits or exams, and prescriptions. Most patients reported to a health care unit devoted to their specific disorder or disease and scheduled follow-up visits every 6 months on average. In this context, potential clients believed that *Abilita* could successfully respond to the following requirements:

- Provision of tools and resources to manage emergency situations (average score of 8.2 on a 0-10 scale, SD 1.6).
- Collection and storage of health care documents and digital contents (average score of 7.7 on a 0-10 scale, SD 3.0).
- Remote communication with authorized health care personnel (average score of 7.6 on a 0-10 scale, SD 2.1).
- Support with monitoring activities (reminders of exams, visits, self-measurements, etc; average score of 6.7 on a 0-10 scale, SD 2.8).
- Targeted information on recreational, informative, or social activities (average score of 6.1 on a 0-10 scale, SD 2.3).

Focus group A identified the Online Help function as a central tool for the app, as it served multiple goals: it accompanies the user in browsing the sections even when he or she has low digital or medical literacy, and it acts as an intermediary between the different users operating within a patient's personal folder.

Focus group B confirmed the main areas of the GUI (menu items) as follows: *Home page; Help; My data; My network; Search; My story; Organizer; Notifications; Personal profile; Info room; Emergency card*. The Online Help, personalized as a female avatar named *Lisa*, interacts with the user by written and/or audiovisual messages. The app also features a medical glossary explaining technical terms and jargon. When users first access the app, *Lisa* provides advice and recommendations on how to start, suggests the sections to be prioritized, and offers easily accessible demos of app functions. In subsequent usage, *Lisa* highlights unread notifications, scheduled appointments, and missing information in the Emergency card when relevant (Figure 1).

The *my data* area is the medical and administrative record and comprises 2 sections: *general outline* and *clinical data and documents* (Figure 2). The sections include *importance or severity* labels that ensure the record's organization and facilitate access to the most relevant data. Key information on the type of disease, therapy, particular care needs, and specific conditions is easily available. Thanks to the *validation* function, HCPs can validate data entered by patients or caregivers.

Figure 1. Home page–shortcuts to the main areas and welcome or follow-up message from Lisa.

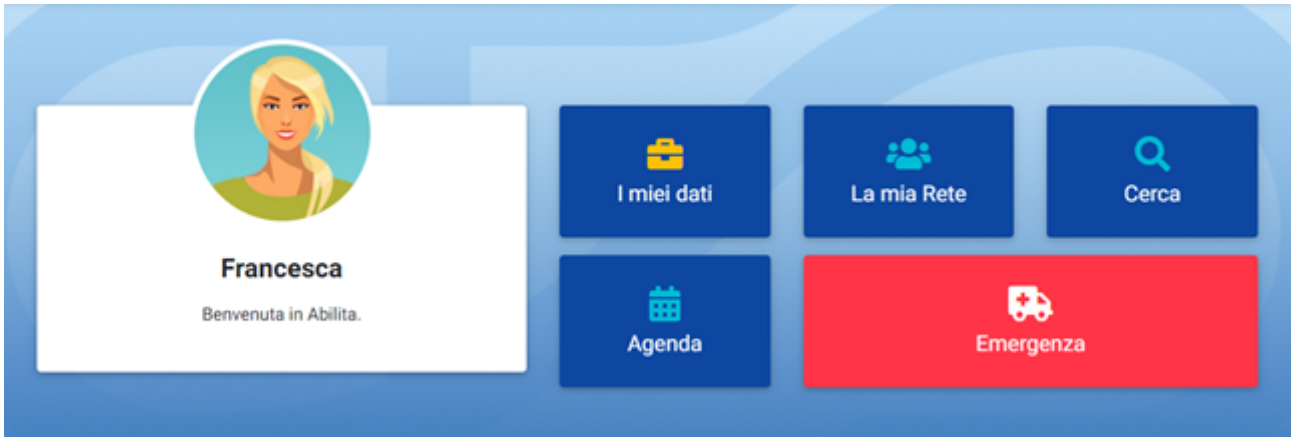
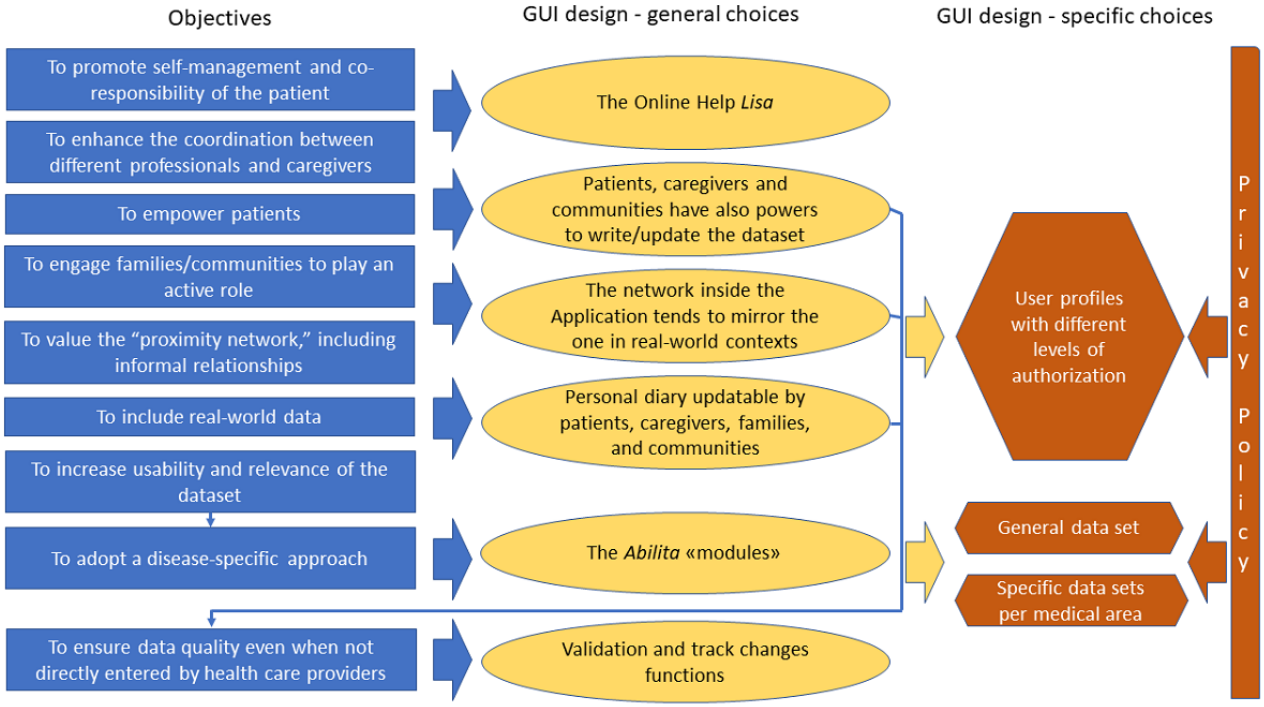


Figure 2. Area “My data.”.

In the area *my network*, the patient or the parent or legal tutor can create a personalized collaborative network of care support (eg, doctors, nurses, parents, friends, neighbors, domestic helpers, babysitters and tutors, teachers, etc). Each member of the network is assigned a separate profile with authorization to access some or all of the personal data. Furthermore, the patient may authorize all health care facilities, thereby enabling all HCP personnel to read and update their medical records. The app also makes available temporary or emergency authorization facilities as well as the blanket withdrawal of all permissions. In the *search* area, it is possible to carry out simple or advanced database searches sorted by data subject or by authorized person (highly recommended by HCPs to facilitate access to relevant information). *My story* hosts a personal diary where users can note clinical data as well as daily experiences, relevant episodes or therapeutic adherence (Multimedia Appendix 1). Actions in the app are always traceable, which allow reconstruction of the author and the date of changes and data validation. Figure 3 summarizes the results of the design process, the relationship between the design and objectives of the research (as discussed in the focus groups) and privacy policy.

In keeping with the privacy policy, the patient is the sole owner and controller of his or her data and the only person able to decide who may treat them and under what conditions, which meets both General Data Protection Regulation requirements and recommendations concerning patient empowerment [22,23]. All sensitive data and interactions between the client (web-based application or emergency mobile app) and the server are encrypted.

Figure 3. The design process.



The results of focus group C confirm that the GUI's disease specificity crucially improves app usability and patient engagement. The relevance of the data set and the perception of utility by families and communities increases when the app is customized based on the specific needs of a subpopulation. In particular, we studied subpopulation management for the following medical areas: autism spectrum disorders, 22q11.2 deletion syndrome, Alzheimer disease, Prader-Willi syndrome, and chronic intestinal failure. The main gaps were centered around the coordination of social and health care services (mostly during follow-up) as well as family support. As a result, the design of the *Abilita modules* for each medical area includes specific GUI features: personalization of the content and structure of the medical data set, contents of the *info room* (information about the disease), and functions of the *organizer* and *notifications* as well as recommendations and priority highlights from Lisa. More specifically, the study foregrounded the following elements:

- Each subpopulation would like to have a personalized page in the *clinical data* subsection.
- Different diseases and ages need differentiated administrative forms.
- The agenda and remind functions could be implemented for specific situations and connected with local networks.
- Users consider it important that data for clinical research at different levels be available.
- Users consider the latest disease-specific documents and recommendations important, such as the Integrated Care Pathway or best clinical practices.

Observational Study, Feedback, and Validation

Table 2 shows the characteristics of enrolled patients and families as well as their average use of the *Abilita* app over the last 4 to 6 months of study. These data were automatically exported by the system administrators and reflect the actions performed by users within the app, including demographic data entered at registration.

Owing to the characteristics of the investigators (pediatricians), most of the enrolled subjects were children or adolescents, in which case the users of the app were mainly parents or family members. HCPs authorized by patients or parents primarily uploaded clinical data and documents. Patients performed operations such as consultation with clinical data, loading of missing clinical investigations, and writing of individual day-to-day experiences. Each patient authorized an average of approximately 2 persons to access their data, who were usually parents and family members, doctors, nurses, and psychologists. By contrast, caregivers and school operators were considerably less involved. The 868 documents that were uploaded included 18 different subtypes, mainly reports of examinations and clinical investigations. Approximately 35% of the data entries were performed by the patients or their parents from the beginning.

We tested 81 *Abilita* functions, which users could access with different levels of authorization (Multimedia Appendix 2). Q3 involved 51 respondents. Table 3 shows the results of the answers to questions 1 to 16, with average positive scores of 78% (4 or 5).

Table 2. Statistics of use of the study population (N=116).

Parameters	Participants
Males, n (%)	67 (57.8)
Age (years), n (%)	
0-10	67 (57.8)
10-20	28 (24.1)
>20	21 (18.1)
Accesses by patients (n=623), mean (SD)	5.4 (2.3)
Authorizations by patients, n	207
Entered documents, n	868
Entered clinic visits, n	307
Entered exams, n	271
Entered diagnoses, n	155
Entered vaccines, n	348
Entered inputs on importance or severity, n	1040
Authorized parents, n	97
Other authorized family members, n	5
Authorized caregivers, n	3
Authorized HCPs[a], n	32

[a]HCP: health care provider.

Table 3. Answers to questions 1-16, expressed in percentage of Likert scale scores.

Question No.	Question	Scores, n (%)		
		1 or 2	3	4 or 5
1.	Is *Abilita* useful for the orderly archiving of medical documents?	1 (2)	6 (12)	44 (86)
2.	Is *Abilita* useful for the orderly archiving of documents concerning care and assistance?	2 (4)	6 (12)	43 (84)
3.	Is *Abilita* useful for remembering the renewal of some clinical evaluations?	0 (0)	10 (20)	41 (80)
4.	Is *Abilita* useful to having your medical history under control everywhere?	1 (2)	2 (4)	48 (94)
5.	Does *Abilita* allow you to monitor some medical parameters when recommended by the HCPs[a]?	2 (4)	7 (14)	42 (82)
6.	Is *Abilita* useful for recording daily self-measurements (eg, blood pressure)?	10 (20)	9 (18)	32 (62)
7.	Does *Abilita* allow you to share information on healthcare or psycho-educational assistance with various professionals?	3 (6)	8 (16)	40 (78)
8.	Does *Abilita* allow you to receive relevant information in a health emergency away from home?	0 (0)	6 (12)	45 (88)
9.	Does *Abilita* allow you to share health information with HCPs without bringing your complete medical chart with you?	0 (0)	3 (6)	48 (94)
10.	Does *Abilita* help you adhere to drug therapy regimens (with reminders) and track what has actually been taken?	6 (12)	10 (20)	35 (68)
11.	Does *Abilita* help you remember which medical devices to buy or order?	6 (12)	14 (27)	31 (61)
12.	Does *Abilita* help you remember administrative deadlines for requesting disability status or for other socio-healthcare procedures?	5 (10)	13 (25)	33 (65)
13.	Does *Abilita* help you to find a document in your archive quickly using advanced search functions?	3 (6)	8 (16)	40 (68)
14.	Does *Abilita* provide useful information about bureaucratic aspects, scientific research or treatments?	2 (4)	14 (27)	35 (69)

Question No.	Question	Scores, n (%)		
		1 or 2	3	4 or 5
15.	Can *Abilita* support HCPs in drawing up a treatment plan and help you follow it?	2 (4)	10 (20)	39 (76)
16.	Overall were you satisfied with the trial run of *Abilita*?	1 (2)	8 (16)	42 (82)

[a]HCP: health care provider.

Questions 17 and 18 asked users about the areas they would like to see enhanced: the answers covered all the areas suggested, with no specific option prevailing significantly, and the same applies to what functions should be integrated (question 18). Interestingly, the option *ability to set preferred tabs or activities to create shortcuts for most used functions* obtained 37% (19/51) of the responses, suggesting that customization is the best strategy. No relevant issues arose regarding privacy and security (questions 19-20): 57% (29/51) of users had no general problems, 65% (33/51) had no problems entering and classifying data, only 23% (12/51) had problems but overcame them with the Lisa online help or with practice (questions 21-30).

Other open and unstructured optional questions (31-36) yielded good feedback concerning the Lisa web-based help, with 47% (24/51) suggesting further implementation of this tool. Patients and caregivers urged informing family doctors and pediatricians about the app to maximize dissemination. The answers on scientific research and on PDTAs (diagnostic-therapeutic assistance pathways) highlight *Abilita*'s potential for data collection subject to privacy consent, for reconstructing analogies in groups of patients affected by the same disease or disorder, and for patient associations to pursue their institutional goals. In addition, *Abilita*'s effectiveness in facilitating relationships or communication with HCPs and local facilities was positively evaluated, preferably with the support of the region. Furthermore, participants considered that the main strengths of the project were *knowledge of one's own medical history with a click* and the overall philosophy behind the app (Multimedia Appendix 3).

Q4, which included 17 predefined questions and addressed 23 HCPs, produced average positive scores of 72% (4 or 5) in the first 16 items defined by a Likert scale score (Multimedia Appendix 4). In the last open-answer item, asking strengths or weaknesses of the project, the following aspects were highlighted:

- The availability of reports and alerts facilitated communication among HCPs and accelerated diagnostic and care paths.
- Users appreciated the involvement of patients or parents in the data entry of documents, lab results, and parameters, although 6 respondents raised concern about quality.
- Overall, 39% (9/23) of respondents encountered general problems in using *Abilita*, especially in the first weeks, and asked that Online Help tools be implemented.
- Users appreciated the importance or severity labels.

Discussion

Principal Findings

The project used needs assessment to establish the contexts to interface with, showing a prevalence of non–self-sufficient patients—typically infants and older adults—diagnosed at an average age of 0 to 5 or 65 to 80 years and mainly supported by health care units specifically devoted to the disorder or disease, for whom follow-up visits are scheduled on average every 6 months. Basic digital skills and good levels of medical literacy of families were identified as starting points of the design.

A sample of 116 patients participated in the observational study. Each patient authorized an average of 1.8 persons to access his or her data, typically parents and family members, doctors, nurses, and psychologists, with the additional involvement of the communities of other institutions and informal environments, for a total of 253 system users. In approximately 35% of cases, data entry was performed by the patients or their parents from the beginning.

Questionnaire Q3 yielded positive patient feedback on the utility of the app to address some health system challenges mentioned as relevant by WHO [24] and on themes such as delayed reporting of events (WHO challenge 1.2), communication roadblocks, lack of access to information or data, insufficient utilization of data and information (WHO challenges 1.4-1.6), insufficient continuity of care, inadequate supportive supervision (WHO challenges 3.5-3.6), low adherence to treatments, and loss of follow-up (WHO challenges 5.2-5.4).

We received no direct evidence on other challenges mentioned by WHO, such as low health worker motivation (3.4), geographic inaccessibility (5.2), insufficient patient engagement (8.1), or absence of community feedback mechanisms (8.3). Some useful indications do emerge in the interpretation of the answers to the same questionnaire Q3. The app promoted communication and team management among HCPs, health care bodies, and families (question 34) and, in addition, increased end user confidence in their own capacity to provide up-to-date, readily searchable, and clear medical information (question 36). According to answers to questions 33 and 35, *Abilita* can contribute to scientific research and PDTA definition (diagnostic-therapeutic assistance pathways), thereby addressing the lack of population denominator (challenge 1.1) —that is, once used by a larger sample of patients in the same medical area, it can become a tool for further assessment of subpopulation management.

The general choices of the GUI design revealed some advantages:

- The GUI is designed around the patient, who is modeled as the main empowered and motivating actor of the actions necessary to maintain and update the medical record.
- Users are constantly supported by the Online Help (avatar *Lisa*), thus addressing medical and digital literacy issues and patient's commitment in terms of his or her specific role, the main problems that arise while using many ICT products.
- Coordination and management needs can be modeled as pathways and actions recommended by Lisa within the app; they are also addressed by targeted functions (search, calendar, and notification areas).
- Real-world data can be traced and collected to then be reused to advance research on the management of complex chronic conditions.

The issue of data quality, indeed highlighted by 6 of the respondents to the HCP survey, was addressed in the project through the track changes and validation functions. It is worth noting that patients and families are increasingly being required to participate in health monitoring, through daily self-measurement and recording of symptoms or in questionnaires, for diseases such as diabetes, and most recently in the COVID-19 pandemic [25,26]. eHealth market engagement strategies—especially in light of the new patient co-responsibility paradigm—are based on flexibility and customization, with a user-friendly design that makes it possible to communicate with or forward information or data to HCPs [27]. In its adoption of these strategies, *Abilita* is in line with a reframed relationship between active citizens and professionals and is intended as a social innovator in the development of a smart community model with the involvement of the proximity network–the app's core feature.

Although informal or territorial networks were not fully exploited by the users during the observational study, as suggested by the number of authorized user profiles (Table 2), we can hypothesize that this was influenced by the study's short duration and the characteristics of the patients involved, mainly children and teenagers. The lockdown period in Italy and Europe revealed the need to innovate public health systems precisely in this direction, linking them to local support networks (through new professional figures such as community nurses) and moving toward an integrated vision of health care. The role of volunteering and associations in providing support to self-isolated and vulnerable persons has also been highlighted [28,29]. In this context, specific design choices may require further refinement, considering, for example, the addition of other user profiles such as territory medicine physician or volunteer.

The modular structure of *Abilita* allows for the personalization of data sets and functions. It also facilitates far-sighted and sustainable investments owing to the partnership's commercial initiatives, which are aimed at developing new modules (optimal feedback has already been received from relevant stakeholders) and intercepting specific target audiences interested in them. Most importantly, this structure allows the patient to choose one or more application *modules* in the case of different pathologies. In this way, *Abilita* has the added value of comorbidity management that is crucial to complex health care populations.

Usability appears to be significantly improved when the GUI is designed according to patients' mental models and when new media and medical literacy are promoted. Following this principle, the assessment of specific subpopulation needs and the development of personalized GUIs for specific medical areas appears important. Procedures to assess patients' needs were successfully experimented and a replicable methodology was defined.

Limitations

This analysis was limited by the low number of enrolled subjects and its short duration. Data collected during the study period and answers to questionnaire Q3 refer mainly to pediatric populations; more evidence is needed about older adult patients' feedback. In fact, only one quarter of them were adults or seniors, but the app was designed and particularly valid for non–self-sufficient subjects, both children and older adults.

The strategy of modular implementation appears to be the best one, but no module has yet been developed and tested. A complete comparison with other available apps, mainly focused on a single disease, will be relevant once the corresponding modules are developed. Specific GUI design choices need to be refined. Nevertheless, the study shows the versatility of this approach for complex health care populations.

Conclusions

eHealth technology allows better management of complex health care aspects in the follow-up of chronic complex disease patients, but translating the UCD into GUI features of an eHealth app is a difficult task. The decision to use patient self-management and co-responsibility as the basis for an eHealth information system seems to have been successful in enhancing the probability of matching the needs of the target population. Moreover, usability appears to be significantly improved when the GUI is designed according to patients' UCD mental models and when new media and medical literacy are promoted. Its potential applications in an era of greater sociosanitary distancing are certainly of particular interest.

Possible lines of exploitation are as follows:

- Design and develop new *Abilita modules* dedicated to specific clinical areas with particular care needs (not least with automatic data download and information managed by the patient's clinical facility of reference).
- Make *Abilita* an integral part of the automatic distribution of data and dissemination of procedures in the public sector (The Italian National Health Care system is structured by regional area, with disease-specific health care facilities that may be very distant from users).
- Strengthen and expand *Abilita* and the patient association network to share information and solutions to the various problems faced by caregivers on a daily basis.
- Simplify usability as much as possible with the possible introduction of voice command shortcuts.

Acknowledgments

The authors thank all the patients and families who agreed to participate in this study. They also thank Dr Alberto E Tozzi for his useful and timely suggestions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots and description of the Italian graphic user interface.
[PDF File (Adobe PDF File), 877 KB - humanfactors_v8i1e18587_app1.pdf]

Multimedia Appendix 2
Tested Abilita functions.
[PDF File (Adobe PDF File), 424 KB - humanfactors_v8i1e18587_app2.pdf]

Multimedia Appendix 3
Feedback questionnaire—patients.
[PDF File (Adobe PDF File), 2061 KB - humanfactors_v8i1e18587_app3.pdf]

Multimedia Appendix 4
Feedback questionnaire—health care providers.
[PDF File (Adobe PDF File), 281 KB - humanfactors_v8i1e18587_app4.pdf]

References

1. Busse R, Blumel M, Scheller-Kreinsen D, Zentner A. Tackling Chronic Disease in Europe - Strategies, Interventions and Challenges. World Health Organization. 2010. URL: http://www.euro.who.int/__data/assets/pdf_file/0008/96632/E93736.pdf/ [accessed 2020-06-17]
2. World Report on Disability. World Health Organization. 2011. URL: https://www.who.int/disabilities/world_report/2011/report.pdf?ua=1 [accessed 2020-12-21]
3. Petersen CL, Weeks WB, Norin O, Weinstein JN. Development and implementation of a person-centered, technology-enhanced care model for managing chronic conditions: cohort study. JMIR Mhealth Uhealth 2019 Mar 20;7(3):e11082 [FREE Full text] [doi: 10.2196/11082] [Medline: 30892274]
4. Simon TD, Berry J, Feudtner C, Stone BL, Sheng X, Bratton SL, et al. Children with complex chronic conditions in inpatient hospital settings in the United States. Pediatrics 2010 Oct;126(4):647-655 [FREE Full text] [doi: 10.1542/peds.2009-3266] [Medline: 20855394]
5. Zulman DM, Jenchura EC, Cohen DM, Lewis ET, Houston TK, Asch SM. How can ehealth technology address challenges related to multimorbidity? Perspectives from patients with multiple chronic conditions. J Gen Intern Med 2015 Aug;30(8):1063-1070 [FREE Full text] [doi: 10.1007/s11606-015-3222-9] [Medline: 25691239]
6. Huygens MW, Vermeulen J, Swinkels IC, Friele RD, van Schayck OC, de Witte LP. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Jul 8;16:232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
7. 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]
8. Ehealth Action Plan 2012-2020: Innovative Healthcare for the 21st Century. European Commission. 2012 Dec 7. URL: https://ec.europa.eu/digital-single-market/en/news/ehealth-action-plan-2012-2020-innovative-healthcare-21st-century/ [accessed 2020-12-21]
9. Agrawal R, Stille C. Building systems that work for children with complex health care needs: editor's note. Pediatrics 2018 Mar;141(Suppl 3):S192-S194 [FREE Full text] [doi: 10.1542/peds.2017-1284C] [Medline: 29496969]
10. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight covid-19 worldwide: challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
11. Electronic Health Record: National guidelines. Italian Ministry of Health. 2011. URL: https://www.gazzettaufficiale.it/home;jsessionid=6FfNPf66R6bX7woN--Vadg__.ntc-as1-guri2a/ [accessed 2018-01-18]

12. Electronic health record: the accession process for the regions has been completed. Agency for digital Italy. 2019. URL: https://www.agid.gov.it/it/agenzia/stampa-e-comunicazione/notizie/2019/10/28/fascicolo-sanitario-elettronico-completato-il-percorso-adesione-regioni/ [accessed 2020-01-18]
13. Rizvi RF, Marquard J, Hultman GM, Adam TJ, Harder KA, Melton GB. Usability evaluation of electronic health record system around clinical notes usage-an ethnographic study. Appl Clin Inform 2017 Oct;8(4):1095-1105 [FREE Full text] [doi: 10.4338/ACI-2017-04-RA-0067] [Medline: 29241247]
14. Guidelines on Hand Hygiene in Health Care. World Health Organization. 2009. URL: https://apps.who.int/iris/bitstream/handle/10665/44102/9789241597906_eng.pdf;jsessionid=3D3FE3A3F185219F1421F81E86955B16?sequence=1 [accessed 2020-12-21]
15. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
16. Reeves D, Blickem C, Vassilev I, Brooks H, Kennedy A, Richardson G, et al. The contribution of social networks to the health and self-management of patients with long-term conditions: a longitudinal study. PLoS One 2014 Jun 2;The Contribution of Social Networks to the Health and Self-Management of Patients with Long-Term Conditions:e98340. [doi: 10.1371/journal.pone.0098340PMID:]
17. Howick J, Kelly P, Kelly M. Establishing a causal link between social relationships and health using the Bradford Hill Guidelines. SSM Popul Health 2019 Aug;8:100402 [FREE Full text] [doi: 10.1016/j.ssmph.2019.100402] [Medline: 31193417]
18. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, 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]
19. Thomas DR. A general inductive approach for analyzing qualitative evaluation data. Am J Eval 2006;27(2):237-246 [FREE Full text]
20. Boncori L. Test Theory and Techniques (Italian). Turin: Bollati Boringhieri; 1993.
21. 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]
22. EUR-Lex: EU law. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/ [accessed 2020-12-21]
23. Ratwani R, Fairbanks T, Savage E, Adams K, Wittie M, Boone E, et al. Mind the Gap. Appl Clin Inform 2017 Dec 18;7(4):1069-1087 [FREE Full text] [doi: 10.4338/aci-2016-06-r-0105]
24. Classification of Digital Health Interventions V1.0. A Shared Language to Describe the Uses of Digital Technology for Health. World Health Organization. 2018. URL: https://www.who.int/reproductivehealth/publications/mhealth/classification-digital-health-interventions/en/ [accessed 2020-12-21]
25. Larbi D, Randine P, Årsand E, Antypas K, Bradway M, Gabarron E. Methods and evaluation criteria for apps and digital interventions for diabetes self-management: systematic review. J Med Internet Res 2020 Jul 6;22(7):e18480 [FREE Full text] [doi: 10.2196/18480] [Medline: 32628125]
26. European mhealth hub. URL: https://mhealth-hub.org/mhealth-solutions-against-covid-19 [accessed 2020-09-18]
27. Alenazi HA, Jamal A, Batais MA. Identification of type 2 diabetes management mobile app features and engagement strategies: modified delphi approach. JMIR Mhealth Uhealth 2020 Sep 11;8(9):e17083 [FREE Full text] [doi: 10.2196/17083] [Medline: 32678798]
28. Rights in the Time of Covid-19 — Lessons From HIV for an Effective, Community-Led Response. Joint United Nations Programme on HIV/AIDS. Switzerland; 2020. URL: https://www.unaids.org/en/resources/documents/2020/human-rights-and-covid-19 [accessed 2020-12-21]
29. Churchill N. Micro-volunteering at scale can help health systems respond to emergencies, such as the Covid-19 pandemic. Patient Experience J 2020 Aug 4;7(2):160-164. [doi: 10.35680/2372-0247.1498]

Abbreviations

GUI: graphic user interface
HCP: health care provider
ICT: information and communications technology
UCD: user-centered design
WHO: World Health Organization

Edited by B Price; submitted 06.03.20; peer-reviewed by B Chaudhry, B Smith; comments to author 27.04.20; revised version received 17.06.20; accepted 03.12.20; published 13.01.21.

Please cite as:
Ferrucci F, Jorio M, Marci S, Bezenchek A, Diella G, Nulli C, Miranda F, Castelli-Gattinara G
A Web-Based Application for Complex Health Care Populations: User-Centered Design Approach
JMIR Hum Factors 2021;8(1):e18587
URL: http://humanfactors.jmir.org/2021/1/e18587/
doi:10.2196/18587
PMID:33439146

Original Paper

Usability, Usefulness, and Acceptance of a Novel, Portable Rehabilitation System (mRehab) Using Smartphone and 3D Printing Technology: Mixed Methods Study

Sutanuka Bhattacharjya[1], MS, PhD, OTR/L; Lora Anne Cavuoto[2], PhD; Brandon Reilly[2]; Wenyao Xu[2], PhD; Heamchand Subryan[2], MArch, MFA; Jeanne Langan[2], PT, PhD

[1]Department of Occupational Therapy, Byrdine F Lewis College of Nursing and Health Professions, Georgia State University, Atlanta, GA, United States

[2]Rehabilitation Science, University at Buffalo, Buffalo, NY, United States

Corresponding Author:
Sutanuka Bhattacharjya, MS, PhD, OTR/L
Department of Occupational Therapy
Byrdine F Lewis College of Nursing and Health Professions
Georgia State University
140 Decatur St SE
Urban Life Building, STE#1246
Atlanta, GA, 30303
United States
Phone: 1 404 413 1449
Email: sutanuka.ot@gmail.com

Abstract

Background: Smart technology use in rehabilitation is growing and can be used remotely to assist clients in self-monitoring their performance. With written home exercise programs being the commonly prescribed form of rehabilitation after discharge, mobile health technology coupled with task-oriented programs can enhance self-management of upper extremity training. In the current study, a rehabilitation system, namely mRehab, was designed that included a smartphone app and 3D-printed household items such as mug, bowl, key, and doorknob embedded with a smartphone. The app interface allowed the user to select rehabilitation activities and receive feedback on the number of activity repetitions completed, time to complete each activity, and quality of movement.

Objective: This study aimed to assess the usability, perceived usefulness, and acceptance of the mRehab system by individuals with stroke and identify the challenges experienced by them when using the system remotely in a home-based setting.

Methods: A mixed-methods approach was used with 11 individuals with chronic stroke. Following training, individuals with stroke used the mRehab system for 6 weeks at home. Each participant completed surveys and engaged in a semistructured interview. Participants' qualitative reports regarding the usability of mRehab were integrated with their survey reports and quantitative performance data.

Results: Of the 11 participants, 10 rated the mRehab system between the 67.5th and 97.5th percentile on the System Usability Scale, indicating their satisfaction with the usability of the system. Participants also provided high ratings of perceived usefulness (mean 5.8, SD 0.9) and perceived ease of use (mean 5.3, SD 1.5) on a 7-point scale based on the Technology Acceptance Model. Common themes reported by participants showed a positive response to mRehab with some suggestions for improvements. Participants reported an interest in activities they perceived to be adequately challenging. Some participants indicated a need for customizing the feedback to be more interpretable. Overall, most participants indicated that they would like to continue using the mRehab system at home.

Conclusions: Assessing usability in the lived environment over a prolonged duration of time is essential to identify the match between the system and users' needs and preferences. While mRehab was well accepted, further customization is desired for a better fit with the end users.

Trial Registration: ClinicalTrials.gov NCT04363944; https://clinicaltrials.gov/ct2/show/NCT04363944

(JMIR Hum Factors 2021;8(1):e21312) doi:10.2196/21312

KEYWORDS

stroke; rehabilitation; smart technology; 3-dimensional printing; usability

Introduction

There are approximately 7 million survivors of stroke in the United States [1]. Up to 60% have residual impairments, which in turn could limit their performance of daily activities [2]. While individuals with stroke are commonly given written home programs when they are discharged from traditional therapies, adherence to written home programs is poor [3]. Qualitative analyses suggest low adherence is related to finding the exercises boring, receiving poor feedback during exercise performance, and uncertainty in how to perform the exercises [3,4]. Mobile health (mHealth) apps provide new options for long-term rehabilitation. In 2018, 91% of adults over the age of 65 years owned a cell phone. Smartphone ownership has increased from 11% in 2011 to 53% in 2018 [5]. As of December 2017, almost 325,000 mHealth apps had been created [6]. However, only a small number of mHealth apps has been specifically designed for people with disabilities, and an even smaller number of apps has undergone accessibility evaluation with people with disabilities [7,8]. Fully assessing usability is critical for the effective and efficient use of mHealth interventions. User feedback on mHealth interventions indicates not all mHealth devices are easy to use [9,10], and this has the potential to limit user adherence. A high dropout rate is one of the most significant barriers to mHealth adoption [11,12]. The average mHealth app costs US $425,000 to develop; however, 83% of mHealth app publishers report a discouraging number of fewer than 10,000 users who activate the app at least once a month [13]. By placing a more significant emphasis on usability for consumers and stakeholders, iterative improvements can reduce costs and enhance the long-term use and adoption of mHealth interventions [14-16]. Thorough usability testing is critical for the success of novel mHealth interventions.

In previous work, a portable system for home rehabilitation, mRehab, was developed and reviewed by end users in a 1-day usability assessment and multiday assessment for consistency in measurement [17]. The system consists of a smartphone and 3D-printed objects in the shapes of household items (a bowl, mug, key, and doorknob; Figure 1). The 3D-printed objects were combined with the smartphone for 10 activities [17,18]. For example, the 3D-printed bowl was designed to hold the smartphone in a landscape orientation. The bowl depth was shallow and had a ridge along the top to allow the user to hold it with both hands (Figure 1). The mug was designed to hold the smartphone in an upright position. Security of the smartphone was ensured by using a screw-top lid on the mug (Figure 2). The mug had a cut-out window for the user to see the smartphone screen during activities. Both left-handed and right-handed mugs were designed. The key and doorknob had similar designs with a pocket holder for the smartphone and mechanical arm that swept across the screen as the object was turned (Figure 3). Two activities, Phone Number and Quick Tap, used the smartphone only and focused on fine motor movements. A wooden box was designed to hold all mRehab items and served as a mechanism to guide participants during horizontal and vertical transfer activities of the bowl or the mug.

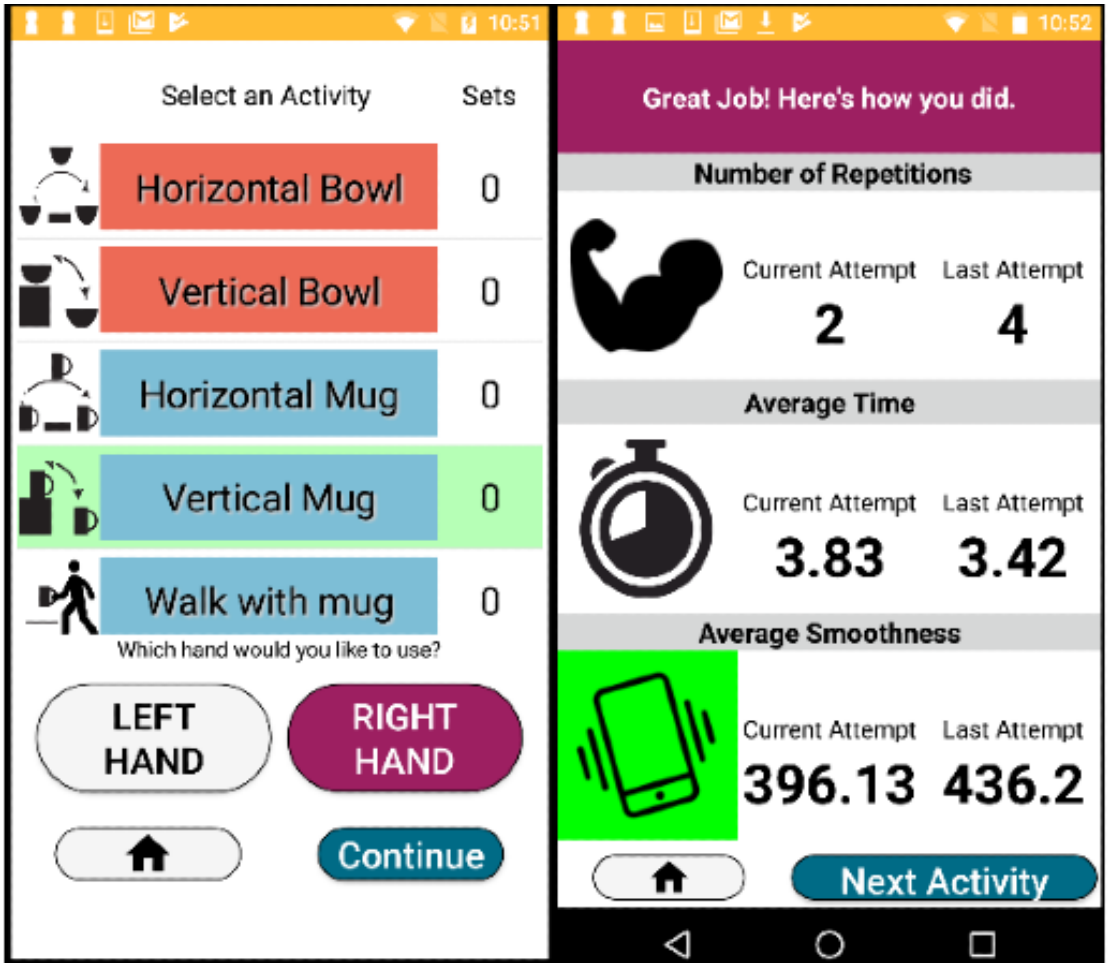
Figure 1. User transferring bowl with both hands.

Figure 2. User seeing feedback on the smartphone screen inside the mug.

Figure 3. User turning doorknob with a smartphone in the holder and the key with a holder.

A Google Nexus 5 phone was used during all mRehab activities. We developed a mobile app that recorded movement-related data (duration and smoothness). This custom app allowed participants to select activities (Figure 4) and then record his or her performance on the activities. Once participants selected an activity, the app provided instructions to guide the user through the activity. A printed manual with instructions was also provided to each user [17]. Additionally, on completion of an activity, the app provided visual feedback in the form of performance scores on the number of repetitions completed, average time to complete a repetition, and average smoothness with which the repetition was completed (Figure 4). The app also provided an auditory readout of the scores. Different from existing technology-based rehabilitation tools, mRehab provides a set of realistic rehabiliaton activities mimicking activities of daily living (ADLs), utilizes a task-oriented approach that focuses on function, and is client-centered. A detailed description of each activity is found in a previous publication [17]. The app also provided performance feedback allowing the user to compare their current performance against their score from the previous session. When the participant's performance (number of repetitions, average time, average smoothness or accuracy) improved over the previous session, the specific icon turned green (eg, average smoothness in Figure 4) and made a celebratory auditory tone to notify the participant they improved [17]. The user could also view a graph that plotted his or her scores from the prior 6 weeks. Previously, we reported on the usability assessment of the previous prototype of mRehab and modifications made that led to the current prototype. We also reported on the consistency of the app measurement for each activity using the current prototype [17].

Figure 4. App interface: activity selection and feedback pages.



In this work, the usability assessment of the system was conducted after a more robust usage of mRehab for 6 weeks at home by 11 individuals with stroke. The examination of usability, usefulness, and acceptance of mRehab holds importance beyond developing this system. Lessons learned about the form and function of mRehab have broad application to mHealth. The use of technology to support home rehabilitation is timely as recommendations to stay home during the COVID-19 pandemic are requiring modifications to health care delivery.

Methods

Research Design

We used a mixed-methods approach, which included quantitative surveys to evaluate long-term usability and perceived usefulness of mRehab, and evaluated the acceptance of the mRehab system. Semistructured interviews with participants were used to further elaborate on the strengths and weaknesses of the mRehab system to better understand the essential ingredients to develop a robust and user-friendly system. The study was approved by the University at Buffalo Institutional Review Board.

Participants

We used a convenience sampling approach to recruit 11 individuals with stroke from the Western New York region who were (1) 18 years of age or older, (2) community dwelling, (3) an independent ambulator, and (4) at least 6 months post stroke. Participants were excluded if any of the following conditions interfered with their participation: (1) cognitive impairment indicated by score of 123 or lower on the Mattis Dementia Scale; (2) acute or chronic pain that would interfere with participation in the study (based upon participant's self-evaluation); (3) severely limited range of motion or contractures of the shoulder, elbow, wrist, or hand that would interfere with participation in the study; (4) absence or severely impaired proprioception of the upper limb; (5) musculoskeletal or circulatory conditions affecting the upper limb; (6) severe spasticity; or (7) recent treatment (within 3 months) for spasticity including botulinum toxin injections or spasticity medications including intrathecal baclofen. Due to a limitation in the number of mRehab units, participants were recruited in 2 rounds: 5 in the first and 6 in the second. All participants provided written informed consent prior to initiating the study.

Procedures

Participants completed 2 in-lab sessions prior to starting the home program. During these sessions, they completed a demographic questionnaire, clinical assessments including the 9-hole peg test and Wolf Motor Function Test, and assessment of hand grip strength and received, in total, 40-60 minutes of training on the mRehab system. In the lab, participants received instructions to select the activity on the mRehab app, insert the smartphone into each 3D-printed object, perform each activity,

and interpret the feedback [17]. Each participant then proceeded to independently complete setting up the mRehab system and perform each activity for 3-5 repetitions to indicate that they were comfortable with setting up and completing the sessions independently. We also explained to the participants that the Quick Twist Mug activity was optional. This activity had lower measurement consistency than we wanted for recommendation in the home program [17], but for those participants willing to use the activity, long-term feedback on performing the activity was considered helpful in furthering the mRehab system. Participant requests for customization such as increasing the font size in the mRehab app for better readability were addressed. For the home program, an occupational therapist suggested that the participants perform 10 repetitions of each activity, 5 times per week as quickly and smoothly as possible. It was clarified that this was only a suggestion and that participants could choose to do more or fewer repetitions. Participants used mRehab at home for 6 weeks and were instructed to contact researchers if they encountered difficulties. After 6 weeks, participants returned to the lab and completed the clinical assessments, showing improved performance [18], and several structured questionnaires. Two questionnaires assessed their general perception towards exercise and technology, the Self-Efficacy for Exercise Scale and the Attitude toward Technology, respectively. The other questionnaires were specific to mRehab: (1) System Usability Scale (SUS); (2) mRehab Acceptance Questionnaire, based on the Technology Acceptance Model; and (3) Difficulty Rating Scale (DRS). Details of each instrument are included in the following sections. Each participant then engaged in a 1-hour retrospective interview conducted by a member of the research team to discuss their experience with using the mRehab system at home. The semistructured interview questions are summarized in Multimedia Appendix 1.

Instruments

Hand Grip Strength Assessment

Hand grip strength assessment using a handheld dynamometer was conducted as part of the Wolf Motor Function Test [19,20]. Hand grip strength assessments were performed for the individuals' affected and nonaffected sides to indicate the individual's baseline motor ability [21].

Self-Efficacy for Exercise (SEE) Scale

On a scale of 1-10, participants indicated their self-efficacy related to exercising in general. Higher scores indicate that participants were more confident that they would complete the exercise when they were alone, stressed, depressed, etc [22].

Attitude Toward Technology

On a scale of 1-7, participants indicated their attitude toward the use of technology in general. Higher scores indicate an increased likelihood that the participant was enthusiastic about using new technology. These questions are based on the Technology Acceptance Model [23-25] and are summarized in Multimedia Appendix 2.

System Usability Scale (SUS)

The SUS has been previously used for assessing usability of mobile rehabilitation apps and systems [26,27]. The SUS consists of 10 questions, each rated on a 5-point Likert scale [28], to assess the participant's satisfaction with the whole mRehab system. The SUS is a reliable and valid measure of the perceived usability of a system [29,30] and has been used with small sample sizes of 8-15 users [31,32]. The SUS was used to assess the participant's satisfaction with the mRehab system.

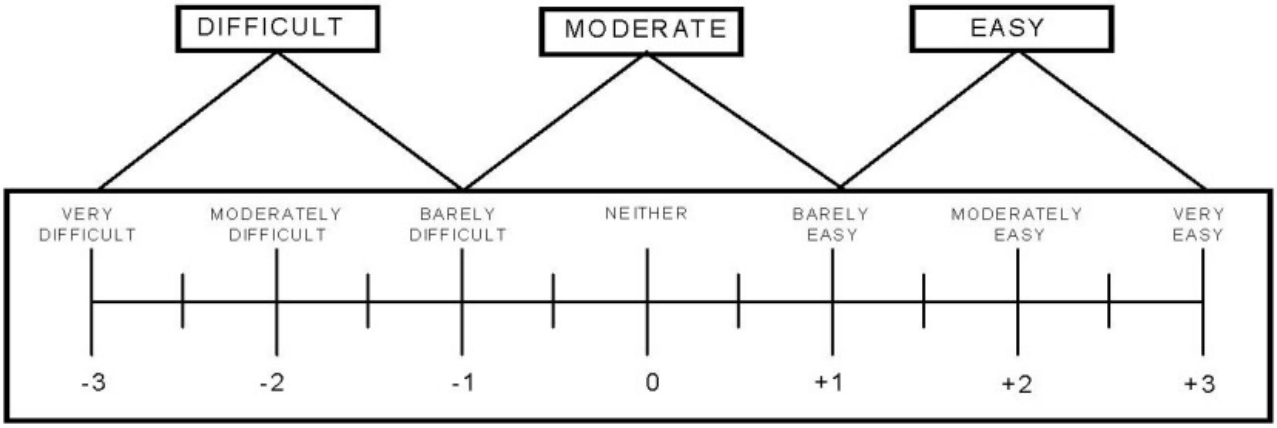
mRehab Acceptance Questionnaire

The mRehab Acceptance Questionnaire was based on the original Technology Acceptance Model and the extended models [33-35]. The questions addressed the mRehab system as a whole and asked about the participant's perception of the system usefulness and ease of use, learnability of the system, self-efficacy for mRehab usage, attitude toward mRehab, and behavioral intention to use the mRehab system in the future. The questions were modified from previous literature [23,36-40] and used a 7-point Strongly Disagree to Strongly Agree Likert-type scale. The questions are summarized in Multimedia Appendix 2.

Difficulty Rating Scale (DRS)

The DRS focused specifically on the hardware design of each of the 3D-printed objects (mug, bowl, key, and doorknob), and elicited participant opinions on their ease of use. Participants rated the ease of use on a scale ranging from Very Difficult to Very Easy (Figure 5).

Figure 5. Ordinal scale on the Difficulty Rating Scale (DRS).



Semistructured Interview

The interview questions elaborated on the usability of the system, including what they liked or disliked about the system components, activities that they benefitted from, and activities that were preferred. Based on the participant responses to the initial probes (see Multimedia Appendix 1), follow-up questions had participants elaborate on their use of the 3D-printed objects and their respective rehabilitation activities.

Data Analysis

Demographic variables are descriptively summarized in Table 1. For the SUS, percentile ranks were calculated from participant ratings of their perceived usability [41]. Grades were assigned to percentile ratings from Grade A to D as recommended by Sauro in 2018 [41]. The assigned cut points for the grades were as follows: A+: 96-100; A: 90-95; A-: 85-89; B+: 80-84; B: 70-79; B-: 65-69; C+: 60-64; C: 41-59; C-: 35-40; and D: 15-34, with grade B- or better indicating acceptable usability and D indicating marginal acceptability. The average of the ratings was calculated for each participant for each subsection of the mRehab Acceptance Questionnaire. Then, the mean and SD were calculated for the mRehab Acceptance Questionnaire for each question across participants [25]. Pearson product moment correlation was used to evaluate the relationship between participants' average number of repetitions performed and their ratings on the SUS and mHealth Acceptance Questionnaire. Use was quantified based on the average number of repetitions per activity over the 6 weeks. Changes in clinical assessments were also examined using the Wolf Motor Function Test and have been reported in another paper [18].

Table 1. Participant characteristics.

ID	Age (years)	Gender	Affected side	Reported dominant arm prior to stroke	Hand grip strength (lb)		Prior experience in using		SEE[a] Scale (1-10 scale), mean	Attitude toward technology (1-7 scale), mean
					Affected side	Nonaffected side	Mobile phone	Smartphone		
S01	57	F[b]	R[c]	R	20	41.7	Y[d]	Y	6.4	2.3
S02	54	F	L[e]	L	25	45	Y	Y	8.2	5.7
S03	68	M[f]	R	R	30	80	Y	Y	10	4.7
S04	61	F	R	R	28.3	41.7	Y	N[g]	6.8	3.3
S05	78	F	L	R	28.3	51.7	Y	N	10	4.7
S06	66	M	L	L	30	111.7	Y	Y	6.9	5
S07	73	M	L	L	10	58.3	Y	N	3.6	3
S08	61	M	L	R	61.7	73.3	Y	Y	6.9	3
S09	62	F	R	R	5	40	Y	Y	6.4	3.3
S10	67	M	R	R	60	60	Y	Y	8.9	2.3
S11	76	M	R	R	45	48.3	Y	N	8.7	2.3
Mean (SD)	65.7 (7.7)	N/A[h]	N/A	N/A	31.2 (18.1)	59.3 (21.8)	N/A	N/A	7.5 (1.9)	3.6 (1.2)

[a]SEE: Self-Efficacy for Exercise.

[b]F: female.

[c]R: right.

[d]Y: yes.

[e]L: left.

[f]M: male.

[g]N: no.

[h]N/A: not applicable.

All interviews were audio recorded and transcribed verbatim by a professional transcription agency. The first author reviewed each transcript for accuracy. QSR's NVivo 12 was then used to code themes within the transcripts. Thematic analysis was used to identify and extract themes, explain what each theme could mean, and determine links between themes. The first author and a research assistant independently coded the transcripts to identify primary and secondary themes from the interview transcripts. Both reviewers discussed their coding once per week over a 6-week coding period and reached mutual consensus in case of any disagreement about coding.

Results

Participant Demographics

The study sample included 11 individuals with stroke, with a mean age of 65.7 (SD 7.7) years and age range of 54-78 years, and 5 of 11 participants were female (46%; detailed in Table 1). On average, the participants were over 7 years poststroke. Of the 11 participants, 8 (73%) were right-side dominant prior to stroke, and 9 (82%) reported that their dominant side was the affected side poststroke. All participants had prior experience with using mobile phones, and most participants (7 out of 11) had prior experience with using a smartphone. On the Attitude Toward Technology, participants reported a mean score of 3.6 (1.2) on the 7-point Strongly Disagree to Strongly Agree scale. All but one participant indicated high self-efficacy for exercise, ranging between 6.4 to 10 in general.

Participant Completion

All but one participant completed the 6-week in-home rehabilitation program. While the participant did not complete the in-home program, they did complete the postintervention interview and all the questionnaires. During the interview, the participant explained that she needed her caregiver to be present during the mRehab sessions. She had difficulty with setting up the mRehab activities and needed support. To better understand this participant's experiences with mRehab, her ratings were included in all reported results.

Issues With the mRehab System

During the in-home period, 6 participants (4 from the first group and 2 from the second group) contacted the research team with reports of breakage in the mRehab system. A majority of the participants in the first group experienced breakage of the doorknob (n=4) and the key (n=2). In case of breakage, the 3D-printed items were replaced within 1-2 days. Following the completion of group 1, we upgraded the 3D-printed items with larger infill to make the doorknobs and keys stronger to withstand repetitive use. In group 2, only 2 participants experienced doorknob breakage.

Perceptions of the mRehab System

Table 2 includes individual-level perceptions of the mRehab system. The SUS scores indicate that all but one participant were satisfied with the usability of the mRehab system. Most participant ratings (10/11) ranged from the 67.5th to the 97.5th percentile, which were Grade B- or better. Participants (11/11) also provided favorable responses on the mRehab Acceptance Questionnaire (a 7-point scale), with a mean perceived usefulness of 5.7, mean perceived ease of use of 5.3, and mean self-efficacy for mRehab usage of 6.0. Also, mean ratings for participants' attitudes toward mRehab was 6.3, and participants' behavioral intention to use mRehab in the future was 5.3. Individual questions for each construct in the mRehab Acceptance Questionnaire have been summarized in Multimedia Appendix 2. For the question "Learning to operate the system was easy for me," participants (11/11) provided a mean rating of 6.1. The average total repetitions of all activities combined per day from the mRehab app are also summarized in Table 2. The correlations between average number of repetitions per day and ratings on SUS, mRehab Acceptance Questionnaire, or DRS were small, and none reached an alpha of .05.

Table 2. Participant ratings on the System Usability Scale (SUS) and mRehab Acceptance Questionnaire and their performance with the mRehab system.

ID	SUS (1-10 scale)		Perceived usefulness (1-7 scale)	Perceived ease of use (1-7 scale), mean	Average repetitions in 6 weeks for all activities
	Percentile	Grade			
S01	17.5	D	5	1.2	N/A[a]
S02	97.5	A+	7	6	189
S03	87.5	A-	7	6.4	255.8
S04	85	A-	6	4.4	256.8
S05	65	B-	5	6.2	461.1
S06	67.5	B-	6	5.8	62.3
S07	80	B+	6	5.2	216
S08	82.5	B+	4	4.6	132.7
S09	80	B+	6	5.4	195.3
S10	67.5	B-	5	6	106.2
S11	95	A	6	5.8	461.5

[a]N/A: not available because the participant did not complete the study.

Participant responses on the DRS indicated that the majority of participants found the mug and bowl easy to use. On the DRS, 7 participants found the mug easy to use, and 4 found it moderately easy to use; 8 participants found the bowl easy to use, and 3 found it moderately easy to use. However, more difficulty was reported with the ease of use of the key and the doorknob. For the doorknob, 3 participants reported it easy to use, 3 reported it as moderate, and 5 indicated it was difficult

to use. For the key, 5 participants reported it easy to use, 3 reported it as moderate, and 3 indicated it was difficult to use.

Themes

The discussion themes identified from the participant interviews are summarized in the following sections: usability of the mRehab system, usability of the performance-based feedback system, usefulness of mRehab activities, support needed with use of mRehab, and generalization to new activities of daily living. The frequency of participant responses reported in the qualitative results represents the number out of all 11 participants.

Usability of the mRehab System

Hardware Design

Comments about the design of the 3D-printed objects were largely positive. Of the 11 participants, 9 liked the bowl, 8 found the mug "good," and 5 liked the doorknob. Comments regarding the design of the doorknob included "an excellent design" and "it was easy to get ahold of it." Regarding the key, 6 participants said that although the key size was bigger than a typical key, they preferred the bigger size for training. The current shape and size allowed a good grip on the key when turning. Some participants pointed out that they would prefer customization of the bowl and mug handle based on the participant's hand size and potentially adding a textured grip on the handle. And 2 participants suggested using a latch or a handle-lever shaped doorknob in the future.

Hardware Functioning

When using the mRehab system at home, 8 participants reported leaving the system set up on a table. Participants thought that the bowl was easy to use during exercise. No difficulties were reported by participants on how to use the mug for the mRehab activities. Regarding using the mug, 5 participants stated that they found the mug easy to use and that the phone was easily accessible when inside the mug. Two participants reported repeated breakage of the doorknob, which led to lower average repetitions for the Turn Doorknob activity. The first 7 participants reported that the doorknob design prevented continuous pairing of the contact interface between the doorknob with the smartphone screen. Some of these participants reported being worried that this could lead to erroneous calculation of smoothness and therefore actively fixed the issue by either placing rolled up paper napkins or a pillbox behind the phone (Figure 6). Additionally, the research team made home visits to attach a piece of foam on the box that pushed the smartphone forward and minimized the space between the smartphone and doorknob, thereby fixing this issue. Since the design of the key was similar to that of the doorknob, there was a similar problem. For 8 participants, initially the app did not register the movement of the key on the phone screen. Again, using an object to push the phone forward toward the key worked well.

Figure 6. Participant using a pill box behind phone when engaging in Turn the Key activity.

Software Design

All participants switched the phone off to preserve battery. Participants reported that the design of the app interface needed to be refined to allow them to make choices on the screen while the phone is in the key or doorknob holder. Two participants reported being pleased by the customizable nature of the app that allowed them to view larger fonts on the screen.

Software Functioning

Two participants reported being confused by the repetition count by the app when they engaged in activities. Participants thought that the app count was directive and they were expected to perform a repetition after the app had counted. The participants reported that they had forgotten that the app counted only after they had completed a repetition. Also, the app count had a brief time lag in counting, which some participants reported to be confusing.

Usability of the Performance-Based Feedback System

Difficulties With the Feedback

Of the 11 participants, 5 participants stated that they did not understand the numbers on the feedback screen and that scores that went to 3 decimal places were not meaningful. One participant explained that they forgot the significance of the auditory celebratory sound and an icon turning green on the feedback screen:

> *I really didn't know [laughs] what I was supposed to be doing–what improvement was. Each time I tried to do them. I was trying to do them as smoothly as I could, and then I was trying to do them all.*

This participant also reported forgetting to look at the manual for a description of the feedback. Although the app was designed to allow participants to see the history of their performance as a line graph over the 6-week period, all participants who remembered the "History" tab (9/11) reported that the app crashed consistently when the history tab was opened. Two participants forgot that the app had a "History" tab and did not remember to look at the manual for more details.

Positives About the Feedback

Of the 11 participants, 9 participants said that they liked the green light and the auditory note of the feedback. One of these participants explained that she deliberately performed 2 sets of each activity everyday with at least one additional repetition in the second set. Performing one extra repetition compared to the previous set ensured that her feedback had at least one green icon for repetitions. One participant explained that the green icon let them identify the activities in which they were becoming "proficient." Another participant said:

> *I liked it when it gave you stats like how well you did, the green light, saying, "Woo! Strong!" that you're getting stronger there and increasing the repetitions. I like the noises that it made.*

One participant said that they tried to redo the activities to get a green icon.

Suggestions for Feedback

Several participants offered suggestions to improve the feedback system; 4 participants said that seeing or hearing the feedback in words could be helpful such as "Today you did faster than yesterday." One participant explained that he would prefer to know what the app was measuring and how he could improve his performance. One participant pointed out that in the activity Walk with Mug, the phone made an initial spilling sound and then stopped. A continuous spilling sound would help.

Two participants said that they would like to see negative feedback. One participant's caregiver explained that the negative feedback could motivate the participant to try another set. One participant requested to include an option to see best score since start. She said:

> *I did it a lot. It got lost. I couldn't tell you what my best score was.*

Usefulness of mRehab Activities

Beneficial Activities

Of the 11 participants, 10 participants reported that they benefitted in some way from one or several of the mRehab activities. Some participants selected more than one activity. Phone Number, Transfer Mug Vertically, and Slow Pour were reported as beneficial by 3 participants. One participant explained that the Slow Pour activity was beneficial for her because it resembled a real-life task. Another participant explained that the horizontal and vertical mug activities were beneficial for her and said, "I can feel it in my shoulder." Phone Number and Quick Tap were reported as beneficial by 2 participants because they required fine motor skills and helped to improved hand-eye coordination. Quick Twist Mug and Transfer Bowl Vertically were not reported as beneficial by any of the participants. Further detail was not provided by 4 participants who reported benefitting from an activity.

Favorable Activity

One or more favorable activities were reported by 10 participants. The only activity not mentioned as a favorite was Turn Doorknob, and the activity mentioned the most, by 5 different participants, was the Transfer Mug Horizontally. The participants did not explain why they enjoyed the activities; they just stated that they liked certain activities more than others.

Nonbeneficial Activities

Eight participants reported not using the Quick Twist Mug activity at all. One of these participants explained that for Quick Twist Mug, the app needed her to quickly supinate and pronate her forearm, and her movement was not quick enough for the app to count the repetition. Walk with Mug and the Transfer Mug Vertically were chosen by 2 participants as nonbeneficial. Turn Doorknob, Turn Key, and Transfer Mug Horizontally were mentioned as nonbeneficial only once. Three participants said that some activities were not beneficial since they were too easy, or they were already able to perform the action with ease before starting the mRehab program.

Nonfavorable Activities

The 4 nonfavorable activities were Slow Pour, Quick Tap, Sip from Mug, and Walk with Mug. Slow Pour was identified as the least favorable activity by 4 participants; 2 of these participants explained they did not like Slow Pour because it forced them to move slowly and they wanted to move faster.

Support Needed to Use mRehab

Four participants indicated that their caregiver helped when using the mRehab system; 3 participants reported needing help with navigating the app, and 1 of the participants felt they could have used the app independently, but defaulted use of the app

to the caregiver because they were more familiar with smartphones. All 4 participants needed physical support with setting up the mRehab activity components. This ranged from assistance with lifting the box to physical assistance with setting up activities. One of the participants indicated that going through all the mRehab activities would take 40-45 minutes and that it was difficult to find free time where their caregiver was available to sit down and help for the entire time.

Generalization to New Daily Life Activities

Nine participants reported initiating a new skill following use of the mRehab system, and 9 participants described an increase in control and use of their affected upper extremity or hand post-mRehab activities. Various ADL performances were brought up by participants: pouring laundry detergent, washing dishes, drying dishes, wiping off countertops, stabilizing with the affected hand, donning socks, opening doorknobs, taking clothes out of dryer, and gripping objects more often. Two participants reported an increase in dexterity of their affected hand post-mRehab activities. Four participants said they were more conscious of using the affected hand during ADLs to continue practicing using it, even outside mRehab activities. Two participants said they did not start doing any new activities, and 1 said it was because they were still experiencing residual pain in their affected hand from their stroke.

Discussion

All participants, except for 1 participant, completed the 6-week study. Overall, participants indicated that they liked using the mRehab system at home and that they benefitted from its use. High percentile ranks on the SUS and high mean ratings on the mRehab Acceptance Questionnaire indicate that the mRehab system was useful as a remote home program and that participants were satisfied with the usability of the system. Although it is possible that individuals who were comfortable with the use of technology volunteered to participate in this study, low scores on the Attitude Toward Technology indicate that the recruited participants were, in general, typically hesitant to try out new technology.

For this study, the inclusion and exclusion criteria were created to ensure that individuals had sufficient function to interact with the system. The criteria, however, did not create a ceiling for the participants. The degree of deficits for individuals varied in the study [18]. By virtue of participants requesting to be in the study, it indicates that they perceive deficits that they would like to improve with a home program. Mild stroke is not uncommon [42], and providing avenues for motor improvement is also important for this group.

The convergence of the qualitative and quantitative data supports the strengths of using a mixed-methods design for capturing a holistic picture for system usability [43]. Participants' ease-of-use ratings and their interview responses indicate that the usability of the mRehab system was high. Participants who described that the bowl and doorknob were easy to use in their interviews also rated them to be +1.5 or higher on the DRS, indicating that they were easy to use. Similarly, participants who described that the design of the 3D-printed key needed to be customized or modified for ease of use rated the key to be moderate to difficult to use on the DRS.

Evaluation of usability over a longer period of time is critical because it portrays the challenges of using a system in day-to-day life while accounting for breakdowns and failures from repeated use. Participants experienced some breakage of the 3D-printed items resulting from repeated and prolonged use. Although the 3D items in the mRehab system had undergone usability testing and were modified based on participant feedback [17], extended use uncovered aspects of the mRehab system that can be improved and expanded in future developments. Participants emphasized the need for customizing the daily use objects in the mRehab system. Also, interviews with the participants revealed technical problems with the "History" tab, which was a newly added feature that was not pilot tested in previous iterations. Despite these issues, the majority of participants provided a grade of A- or better for mRehab on the SUS. Scores that are 68th percentile or higher on the SUS suggests future use of the system [16,28,29]. Both the perceived usefulness and perceived ease-of-use scores suggested the participants were satisfied and were accepting of the mRehab system. The Technology Acceptance Model posits that perceived usefulness and perceived ease of use are 2 main factors that predict actual use of the technology by the user and influence acceptance [23,44].

Although participants reported quickly learning to use the system in the mRehab Acceptance Questionnaire, the interviews revealed that they did not have a full understanding of the app interface or the feedback system. Over the 6-week period, participants had forgotten what the scores (numbers) meant, what the visual feedback (green light) was, and what the celebratory auditory note meant. These behaviors indicate that 40-60 minutes of training was not adequate for the participants to use the system to its fullest capacity in a remote setting. Relatedly, hospital-based research suggests transition planning and early training prior to discharge from hospital are important to facilitate carry over of skills to remote rehabilitation and promote self-management [45]. All participants had received a manual explaining the meaning and significance of each activity and the app interface; however, the participants reported either forgetting about the manual or not taking it out of the box. This indicates that the participants relied on the app to guide them through the entire exercise session. Better understanding how to support individuals in long-term home programs through in-person training and app design are important considerations for design and implementation of mHealth.

The long-term use of mRehab combined with multiple assessments of usability testing start to illuminate the individual's preference for activities that are just right and are neither too easy nor too difficult. Participants' preferences for the just-right amount of challenge have been demonstrated in previous literature [46,47]. Participants explained that they did not benefit from activities that were too easy. Conversely, several participants stopped using the Quick Twist Mug activity because it was too challenging. Also, with the Slow Pour activity, participants listed it as "not a favorite," but reported they did the activity and found it beneficial. Taken together, it suggests that feeling appropriately challenged and benefiting

from an activity are important aspects to consider in designing rehabilitation systems.

This was a small-scale, mixed-methods study to explore the feasibility of using mHealth relatively independently for upper limb rehabilitation by individuals with stroke. This sample size may not have allowed us to identify all the possible accessibility features needed by people with disabilities, but the in-depth conversations with these study participants enabled us to identify several major accessibility features desired by individuals with stroke. Additionally, despite immediate replacements, the breakage of some of the 3D-printed items may have caused negative perceptions about the mRehab system. However, the participants provided an overall positive usability rating for the mRehab system. The first group of participants experienced a higher incidence of breakage than the second group. Although our plan did not entail using an iterative approach within this study, the first group's home use of the 3D-printed items allowed us to modify the 3D-printed objects for the second group. The benefits of extended use of a device prior to usability testing are well illustrated in this study.

During screening, participants were included if they indicated in their self-assessment that pain would not interfere with their participation. Experiencing pain is a common clinical consequence after stroke [48], and nearly 70% of poststroke patients experience pain on a daily basis [49]. Postintervention, 2 participants reported not engaging in new activities, fearing pain. The usability assessments in this study did not fully evaluate if mRehab activities resulted in pain. At the start of the study, participants were instructed to stop mRehab activities if they experienced increased pain and to contact the research team. No participant contacted the research team with complaints of pain. Perceived fear of pain when performing a new activity may also impact the participant's willingness to engage in new activities. In previous studies, participants reported planning daily activities with their nonaffected side due to fear of injury to their affected arm [50,51]. In future studies, a pain scale on the mobile app that records reports of pain and assessing fear of pain with movement will help clarify how pain and the fear of pain impact outcomes. This line of study is important in better understanding how training in rehabilitation programs may transfer to movement outside of the rehabilitation program.

Assessing usability and usefulness of mHealth interventions is critical to incorporate user opinions and customize the intervention to the users' needs and preferences. It is not common for end users to evaluate their exercises [52], let alone assess long-term usability in the user's lived environment. Findings from this study indicated users' preferences for (1) realistic design of the 3D-printed objects, (2) activities resembling daily living tasks, (3) customizable nature of the app, (4) being adequately challenged by the activities, and (5) performance-based objective auditory and visual feedback.

Acknowledgments

This work was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number R21 HD092243-01.

Authors' Contributions

SB wrote the first draft of the manuscript. JL and SB oversaw recruitment and training. WX oversaw app function and mRehab data collection. SB and BR conducted the qualitative analysis. LC conducted the quantitative analysis. LC and HS oversaw mRehab system design. All authors provided suggestions or revisions to the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Format of the semi-structured interview.
[DOCX File , 15 KB - humanfactors_v8i1e21312_app1.docx]

Multimedia Appendix 2
Questions based on Technology Acceptance Model.
[DOCX File , 24 KB - humanfactors_v8i1e21312_app2.docx]

References

1. Benjamin EJ, Blaha MJ, Chiuve SE, Cushman M, Das SR, Deo R, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. Circulation 2017 Mar 07;135(10):e146-e603 [FREE Full text] [doi: 10.1161/CIR.0000000000000485] [Medline: 28122885]
2. Doman CA, Waddell KJ, Bailey RR, Moore JL, Lang CE. Changes in Upper-Extremity Functional Capacity and Daily Performance During Outpatient Occupational Therapy for People With Stroke. Am J Occup Ther 2016;70(3):7003290040p1-7003290040p11 [FREE Full text] [doi: 10.5014/ajot.2016.020891] [Medline: 27089298]

3. Argent R, Daly A, Caulfield B. Patient Involvement With Home-Based Exercise Programs: Can Connected Health Interventions Influence Adherence? JMIR Mhealth Uhealth 2018 Mar 01;6(3):e47 [FREE Full text] [doi: 10.2196/mhealth.8518] [Medline: 29496655]
4. Novak I. Effective home programme intervention for adults: a systematic review. Clin Rehabil 2011 Dec;25(12):1066-1085. [doi: 10.1177/0269215511410727] [Medline: 21831927]
5. Mobile Fact Sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-03-01]
6. mHealth Economics 2017 – Current Status and Future Trends in Mobile Health. Research2Guidance. 2017 Nov. URL: https://research2guidance.com/product/mhealth-economics-2017-current-status-and-future-trends-in-mobile-health/ [accessed 2021-03-01]
7. DeRuyter F, Jones M. Mobile Healthcare and mHealth Apps for People with Disabilities. In: Miesenberger K, Kouroupetroglou G, editors. Computers Helping People with Special Needs. ICCHP 2018. Lecture Notes in Computer Science, vol 10897. Cham, Switzerland: Springer Publishing Company; 2018.
8. DeRuyter F, Jones ML, Morris JT. Mobile Health Apps and Needs of People with Disabilities: A National Survey. Journal on Technology & Persons with Disabilities 2018;6:161 [FREE Full text]
9. Hornbæk K. Current practice in measuring usability: Challenges to usability studies and research. International Journal of Human-Computer Studies 2006 Feb;64(2):79-102 [FREE Full text] [doi: 10.1016/j.ijhcs.2005.06.002]
10. 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]
11. Dorsey ER, Yvonne Chan YF, McConnell MV, Shaw SY, Trister AD, Friend SH. The Use of Smartphones for Health Research. Acad Med 2017 Feb;92(2):157-160. [doi: 10.1097/ACM.0000000000001205] [Medline: 27119325]
12. Mayberry LS, Mulvaney SA, Johnson KB, Osborn CY. The MEssaging for Diabetes Intervention Reduced Barriers to Medication Adherence Among Low-Income, Diverse Adults With Type 2. J Diabetes Sci Technol 2017 Jan;11(1):92-99 [FREE Full text] [doi: 10.1177/1932296816668374] [Medline: 27595710]
13. mHealth Economics - How mHealth App Publishers Are Monetizing Their Apps. Research2Guidance. 2018 Mar. URL: https://research2guidance.com/product/mHealth-economics-how-mHealth-app-publishers-are-monetizing-their-apps/ [accessed 2021-03-01]
14. Ribeiro N, Moreira L, Barros A, Almeida AM, Santos-Silva F. Guidelines for a cancer prevention smartphone application: A mixed-methods study. Int J Med Inform 2016 Oct;94:134-142. [doi: 10.1016/j.ijmedinf.2016.07.007] [Medline: 27573321]
15. Smith SA, Claridy MD, Whitehead MS, Sheats JQ, Yoo W, Alema-Mensah EA, et al. Lifestyle Modification Experiences of African American Breast Cancer Survivors: A Needs Assessment. JMIR Cancer 2015;1(2) [FREE Full text] [doi: 10.2196/cancer.4892] [Medline: 26380378]
16. Benefits of User-Centered Design. US General Services Administration. 2021. URL: https://www.usability.gov/what-and-why/benefits-of-ucd.html [accessed 2021-03-01]
17. Bhattacharjya S, Stafford MC, Cavuoto LA, Yang Z, Song C, Subryan H, et al. Harnessing smartphone technology and three dimensional printing to create a mobile rehabilitation system, mRehab: assessment of usability and consistency in measurement. J Neuroeng Rehabil 2019 Oct 29;16(1):127 [FREE Full text] [doi: 10.1186/s12984-019-0592-y] [Medline: 31665036]
18. Langan J, Bhattacharjya S, Subryan H, Xu W, Chen B, Li Z, et al. In-Home Rehabilitation Using a Smartphone App Coupled With 3D Printed Functional Objects: Single-Subject Design Study. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e19582 [FREE Full text] [doi: 10.2196/19582] [Medline: 32706702]
19. Wolf SL, Catlin PA, Ellis M, Archer AL, Morgan B, Piacentino A. Assessing Wolf motor function test as outcome measure for research in patients after stroke. Stroke 2001 Jul;32(7):1635-1639. [doi: 10.1161/01.str.32.7.1635] [Medline: 11441212]
20. Wolf SL, Catlin PA, Ellis M, Link A, Morgan B, Piacento A. Wolf Motor Function Test (WMFT). Measurement Instrument Database for the Social Sciences (MIDSS). 2021. URL: https://www.midss.org/wolf-motor-function-test-wmft [accessed 2021-03-01]
21. Ekstrand E, Lexell J, Brogårdh C. Grip strength is a representative measure of muscle weakness in the upper extremity after stroke. Top Stroke Rehabil 2016 Dec;23(6):400-405. [doi: 10.1080/10749357.2016.1168591] [Medline: 27145212]
22. Resnick B, Jenkins LS. Testing the reliability and validity of the Self-Efficacy for Exercise scale. Nurs Res 2000;49(3):154-159. [doi: 10.1097/00006199-200005000-00007] [Medline: 10882320]
23. Alharbi S, Drew S. Using the Technology Acceptance Model in Understanding Academics' Behavioural Intention to Use Learning Management Systems. IJACSA 2014;5(1). [doi: 10.14569/ijacsa.2014.050120]
24. Jeffrey DA. Testing the technology acceptance model 3 (TAM 3) with the inclusion of change fatigue and overload, in the context of faculty from Seventh-day Adventist universities: A revised model. Dissertations. 2015. URL: https://digitalcommons.andrews.edu/dissertations/1581 [accessed 2021-03-01]
25. Park SY. An Analysis of the Technology Acceptance Model in Understanding University Students' Behavioral Intention to Use e-Learning. Journal of Educational Technology & Society 2009;12(3):150-162 [FREE Full text]

26. Meldrum D, Glennon A, Herdman S, Murray D, McConn-Walsh R. Virtual reality rehabilitation of balance: assessment of the usability of the Nintendo Wii(®) Fit Plus. Disabil Rehabil Assist Technol 2012 May;7(3):205-210. [doi: 10.3109/17483107.2011.616922] [Medline: 22117107]
27. Vanbellingen T, Filius SJ, Nyffeler T, van Wegen EEH. Usability of Videogame-Based Dexterity Training in the Early Rehabilitation Phase of Stroke Patients: A Pilot Study. Front Neurol 2017;8:654 [FREE Full text] [doi: 10.3389/fneur.2017.00654] [Medline: 29276499]
28. Sauro J. Measuring usability with the system usability scale (SUS). MeasuringU. 2011 Feb 2. URL: https://measuringu.com/sus/ [accessed 2021-02-12]
29. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594 [FREE Full text] [doi: 10.1080/10447310802205776]
30. 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. Berlin, Germany: Springer Publishing Company; 2009.
31. Nehrujee A, Vasanthan L, Lepcha A, Balasubramanian S. A Smartphone-based gaming system for vestibular rehabilitation: A usability study. J Vestib Res 2019;29(2-3):147-160. [doi: 10.3233/VES-190660] [Medline: 31356221]
32. Pedroli E, Greci L, Colombo D, Serino S, Cipresso P, Arlati S, et al. Characteristics, Usability, and Users Experience of a System Combining Cognitive and Physical Therapy in a Virtual Environment: Positive Bike. Sensors (Basel) 2018 Jul 19;18(7) [FREE Full text] [doi: 10.3390/s18072343] [Medline: 30029502]
33. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319. [doi: 10.2307/249008]
34. Davis FD, Bagozzi RP, Warshaw PR. User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 1989 Aug;35(8):982-1003 [FREE Full text] [doi: 10.1287/mnsc.35.8.982]
35. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204 [FREE Full text] [doi: 10.1287/mnsc.46.2.186.11926]
36. Brown W, Yen P, 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 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
37. Georgsson M, Staggers N. Patients' Perceptions and Experiences of a mHealth Diabetes Self-management System. Comput Inform Nurs 2017 Mar;35(3):122-130. [doi: 10.1097/CIN.0000000000000296] [Medline: 27748662]
38. 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]
39. Ma Q, Liu L. The Technology Acceptance Model: A Meta-Analysis of Empirical Findings. Journal of Organizational and End User Computing (JOEUC) 2004 Jan;16(1):59-72. [doi: 10.4018/joeuc.2004010104]
40. Yarbrough AK, Smith TB. Technology acceptance among physicians: a new take on TAM. Med Care Res Rev 2007 Dec;64(6):650-672. [doi: 10.1177/1077558707305942] [Medline: 17717378]
41. Sauro J. 5 ways to interpret a SUS score. MeasuringU. 2018 Sep 19. URL: https://measuringu.com/interpret-sus-score/ [accessed 2021-03-01]
42. Edwards DF, Hahn M, Baum C, Dromerick AW. The impact of mild stroke on meaningful activity and life satisfaction. J Stroke Cerebrovasc Dis 2006;15(4):151-157. [doi: 10.1016/j.jstrokecerebrovasdis.2006.04.001] [Medline: 17904068]
43. Carayon P, Kianfar S, Li Y, Xie A, Alyousef B, Wooldridge A. A systematic review of mixed methods research on human factors and ergonomics in health care. Appl Ergon 2015 Nov;51:291-321 [FREE Full text] [doi: 10.1016/j.apergo.2015.06.001] [Medline: 26154228]
44. Kim D, Chang H. Key functional characteristics in designing and operating health information websites for user satisfaction: an application of the extended technology acceptance model. Int J Med Inform 2007 Dec;76(11-12):790-800. [doi: 10.1016/j.ijmedinf.2006.09.001] [Medline: 17049917]
45. Cameron JI, Tsoi C, Marsella A. Optimizing stroke systems of care by enhancing transitions across care environments. Stroke 2008 Sep;39(9):2637-2643. [doi: 10.1161/STROKEAHA.107.501064] [Medline: 18635864]
46. Guadagnoli MA, Lee TD. Challenge point: a framework for conceptualizing the effects of various practice conditions in motor learning. J Mot Behav 2004 Jun;36(2):212-224. [doi: 10.3200/JMBR.36.2.212-224] [Medline: 15130871]
47. Woodbury ML, Anderson K, Finetto C, Fortune A, Dellenbach B, Grattan E, et al. Matching Task Difficulty to Patient Ability During Task Practice Improves Upper Extremity Motor Skill After Stroke: A Proof-of-Concept Study. Arch Phys Med Rehabil 2016 Nov;97(11):1863-1871 [FREE Full text] [doi: 10.1016/j.apmr.2016.03.022] [Medline: 27117385]
48. Harrison RA, Field TS. Post stroke pain: identification, assessment, and therapy. Cerebrovasc Dis 2015;39(3-4):190-201 [FREE Full text] [doi: 10.1159/000375397] [Medline: 25766121]
49. Klit H, Finnerup NB, Andersen G, Jensen TS. Central poststroke pain: a population-based study. Pain 2011 Apr;152(4):818-824. [doi: 10.1016/j.pain.2010.12.030] [Medline: 21272999]
50. Damush TM, Plue L, Bakas T, Schmid A, Williams LS. Barriers and facilitators to exercise among stroke survivors. Rehabil Nurs 2007;32(6):253-60, 262. [doi: 10.1002/j.2048-7940.2007.tb00183.x] [Medline: 18065147]
51. Widar M, Ek A, Ahlström G. Coping with long-term pain after a stroke. J Pain Symptom Manage 2004 Mar;27(3):215-225 [FREE Full text] [doi: 10.1016/j.jpainsymman.2003.07.006] [Medline: 15010100]

52. O'Reilly M, Caulfield B, Ward T, Johnston W, Doherty C. Wearable Inertial Sensor Systems for Lower Limb Exercise Detection and Evaluation: A Systematic Review. Sports Med 2018 May;48(5):1221-1246. [doi: 10.1007/s40279-018-0878-4] [Medline: 29476427]

Abbreviations

ADLs: activities of daily living
DRS: Difficulty Rating Scale
mHealth: mobile health
SEE: Self-Efficacy for Exercise
SUS: System Usability Scale

Edited by A Kushniruk; submitted 11.06.20; peer-reviewed by K Laver, S Hoermann, L Shum; comments to author 03.07.20; revised version received 24.08.20; accepted 19.02.21; published 22.03.21.

Please cite as:
Bhattacharjya S, Cavuoto LA, Reilly B, Xu W, Subryan H, Langan J
Usability, Usefulness, and Acceptance of a Novel, Portable Rehabilitation System (mRehab) Using Smartphone and 3D Printing Technology: Mixed Methods Study
JMIR Hum Factors 2021;8(1):e21312
URL: https://humanfactors.jmir.org/2021/1/e21312
doi:10.2196/21312
PMID:33749608

Review

Procedures of User-Centered Usability Assessment for Digital Solutions: Scoping Review of Reviews Reporting on Digital Solutions Relevant for Older Adults

Anabela G Silva[1], PhD; Hilma Caravau[2], MSci; Ana Martins[2], PhD; Ana Margarida Pisco Almeida[3], PhD; Telmo Silva[3], PhD; Óscar Ribeiro[4], PhD; Gonçalo Santinha[5], PhD; Nelson P Rocha[2], PhD

[1]School of Health Sciences, Center for Health Technology and Services Research, University of Aveiro, Aveiro, Portugal

[2]Department of Medical Sciences, Institute of Electronics and Informatics Engineering of Aveiro, Aveiro, Portugal

[3]DigiMedia, Department of Communication and Art, University of Aveiro, Aveiro, Portugal

[4]Department of Education and Psychology, Center for Health Technology and Services Research, University of Aveiro, Aveiro, Portugal

[5]Department of Social, Political and Territorial Sciences, Governance, Competitiveness and Public Policies, University of Aveiro, Aveiro, Portugal

Corresponding Author:
Anabela G Silva, PhD
School of Health Sciences
Center for Health Technology and Services Research
University of Aveiro
Campus Universitário de Santiago
Aveiro, 3810-193
Portugal
Phone: 351 234401558
Email: asilva@ua.pt

Abstract

Background: The assessment of usability is a complex process that involves several steps and procedures. It is important to standardize the evaluation and reporting of usability procedures across studies to guide researchers, facilitate comparisons across studies, and promote high-quality usability studies. The first step to standardizing is to have an overview of how usability study procedures are reported across the literature.

Objective: This scoping review of reviews aims to synthesize the procedures reported for the different steps of the process of conducting a user-centered usability assessment of digital solutions relevant for older adults and to identify potential gaps in the present reporting of procedures. The secondary aim is to identify any principles or frameworks guiding this assessment in view of a standardized approach.

Methods: This is a scoping review of reviews. A 5-stage scoping review methodology was used to identify and describe relevant literature published between 2009 and 2020 as follows: identify the research question, identify relevant studies, select studies for review, chart data from selected literature, and summarize and report results. The research was conducted on 5 electronic databases: PubMed, ACM Digital Library, IEEE, Scopus, and Web of Science. Reviews that met the inclusion criteria (reporting on user-centered usability evaluation procedures for any digital solution that could be relevant for older adults and were published in English) were identified, and data were extracted for further analysis regarding study evaluators, study participants, methods and techniques, tasks, and test environment.

Results: A total of 3958 articles were identified. After a detailed screening, 20 reviews matched the eligibility criteria. The characteristics of the study evaluators and participants and task procedures were only briefly and differently reported. The methods and techniques used for the assessment of usability are the topics that were most commonly and comprehensively reported in the reviews, whereas the test environment was seldom and poorly characterized.

Conclusions: A lack of a detailed description of several steps of the process of assessing usability and no evidence on good practices of performing it suggests that there is a need for a consensus framework on the assessment of user-centered usability evaluation. Such a consensus would inform researchers and allow standardization of procedures, which are likely to result in improved study quality and reporting, increased sensitivity of the usability assessment, and improved comparability across studies

and digital solutions. Our findings also highlight the need to investigate whether different ways of assessing usability are more sensitive than others. These findings need to be considered in light of review limitations.

(JMIR Hum Factors 2021;8(1):e22774) doi:10.2196/22774

KEYWORDS

mobile phone; user-centered design; aged; review; telemedicine

Introduction

Background

Digital solutions, defined as any set of technologies, systems, and mobile apps that are available on a digital device such as an iPad, a laptop, or a smartphone [1], have become popular in different areas, namely to optimize and personalize health care provision [2], to promote healthy lifestyles (eg, physical activity) [3,4], to minimize loneliness and social exclusion by promoting social, religious, civic, and political participation [5-7], or to improve safety, independence, and confidence [2].

The accelerated aging of the population imposes several challenges on the health care and social systems. Owing to the higher rates of disease and morbidity [8,9], digital solutions have been noted as a valid contributor to help reach a high number of individuals at lower costs [10]. However, developing digital solutions adjusted to older adults presents specific challenges related to age and disease, such as loss of visual and hearing acuity or changes in fine motricity. These need to be considered so that the technology matches the users' needs and characteristics and, ultimately, its use results in an added value in daily life [11,12]. To guarantee that a digital solution is fully adjusted to its users, a robust evaluation process must be considered [13]. One of the key attributes of digital solutions that require careful attention and evaluation is usability.

Usability is part of the user experience, that is, the total usage phenomenon [14], and is defined as the measure by which a product can be used by specific users to achieve specific goals with effectiveness, efficiency, and satisfaction in a specific context of use [15]. Efficacy refers to the degree of accuracy and completeness with which users achieve certain goals in a given environment, efficiency is related to the accuracy and completeness of the goals achieved with regard to the resources used, and satisfaction is defined as the comfort and acceptance on the use of a system [15]. Furthermore, the level of usability obtained depends on the specific circumstances in which the product is used and the usage context includes users, tasks, equipment (hardware and software), and the physical and social environment, as all of these factors can influence the usability of digital solutions [15]. In other words, usability is the ability of a product to be understood, learned, used, and attractive to the user, when used under specific conditions. This definition reinforces the idea that a product has no intrinsic usability and only the ability to be used under specific conditions [16]. Good usability allows reducing task execution times, errors, or learning times; improves user satisfaction; and leads to improved product acceptability, increased user satisfaction, and improved product reliability [17].

Usability evaluation is an important part of the overall development of user interaction mechanisms, which consists of interactive cycles of design, prototyping, and validation [18]. Ideally, usability evaluation must be present at all development stages and must be iterative to enable a continuous evolution of the quality of the product or service. The literature describes several models, methods, and techniques to ensure that usability issues are considered during the development process. The selection of these models, methods, and techniques depends on the development stage of digital solutions and available resources [19]. Certain models of usability assessment rely on usability experts, whereas others rely on end users (user-centered usability assessment). The former are known as the analytical models [20] and involve the inspection of the digital solution by experts to assess the various aspects of user interaction against an established set of principles of interface design and usability [21,22]. The latter refer to the empirical models [20] and involve having the perspective of users and are key to highly usable digital solutions by ensuring that the digital solutions meet the users' needs and requirements, that is, they are adapted to the body and mind of their user in a given context [23]. This perspective is gathered using different methods (eg, test and inquiry) and techniques (eg, interviews, think-aloud, and observation), which are usually combined [24]. Both models are essential in the development process of digital solutions and provide complementary information [25]. This review focuses on the users' assessment of usability.

Usability assessment involving users is a complex task, and the use of only one method (eg, test or inquiry) may not be comprehensive enough to thoroughly consider all relevant issues associated with a given product or service [19]. In addition, different methods have different strengths and weaknesses and provide information on different aspects of the digital solution [19]. Nevertheless, it is important to standardize the evaluation and reporting of usability procedures across studies. This will guide researchers, facilitate comparisons across studies, promote high-quality usability studies, which would be more likely to identify usability problems, and provide relevant data that contribute to highly usable solutions. The first step to standardizing is to provide an overview of how user-centered usability evaluation procedures are reported in the literature.

Objective

This scoping review of reviews aims to synthesize the procedures used or reported for the different steps of the process of conducting a user-centered usability assessment of digital solutions relevant for older adults and identify potential gaps in the present reporting of procedures. The secondary aim is to identify the principles guiding this assessment.

Methods

Study Design

This study followed the 5-stage scoping review methodology defined by Levac et al [26] based on the framework previously developed by Arskey and O'Malley [27]. The stages include (1) identification of the research question, (2) identification of relevant studies, (3) selection of relevant studies, (4) charting the data, and (5) collating, summarizing, and reporting the results of the review. A scoping review of the literature aims to map key concepts, summarize a range of evidence, especially in complex fields, and identify gaps in the existing literature. It allows for broader perspectives in comparison with systematic reviews [26,27] and, therefore, was the appropriate approach for this study, in which we aimed to cover a broad range of usability evaluation procedures and identify gaps to direct future research.

Identification of the Research Question

The research question provides a roadmap for the subsequent stages of the review. It was defined based on the analysis of the literature in the field of usability evaluation of digital solutions and the expertise of the research team, that is, during our previous work in the field of usability evaluation, we identified a lack of consensus in the academic literature regarding the instruments, protocols, and methodologies used for assessing usability across a range of digital solutions (eg, websites, assistive technology, augmented reality). Therefore, to have a more in-depth knowledge of the practices and procedures used, the following research question was defined: *What are the current practices for the user-centered assessment of the usability of digital solutions (eg, procedures instruments) relevant (ie, that could be used and have value) for the older adult population?* This broad question was subdivided into 5 research questions: (1) What are the characteristics of study evaluators reported in user-centered usability studies for digital solutions relevant to older adults? (2) What are the characteristics of study participants reported in user-centered usability studies for digital solutions relevant to older adults? (3) How are the tasks used for user-centered usability studies for digital solutions relevant to older adults? (4) What are the methods and techniques used in user-centered usability studies for digital solutions relevant to older adults? and (5) Where (ie, the environment) do user-centered usability evaluations take place?

Identification of Relevant Studies

The search expression *usability* OR *user experience* was used in the electronic search carried out in PubMed, ACM Digital Library, IEEE, Scopus, and Web of Science. The search expression did not include *older adults* as we did not want to limit the inclusion of reviews to those specifically mentioning *older adults*. Databases were searched for English language reviews published between January 1, 2009, and January 23, 2020. The limit of 2009 was established, as 2007 was the year the *ambient assisted living* joint programme was launched by the European Commission, which is a transnational funding program exclusively focused on the research and development of digital solutions directed at older adults [28]. Therefore, we searched for reviews from 2009 onward that covered the primary studies published after 2007.

Selection of Relevant Studies

All references were imported into Mendeley software (Elsevier, North-Holland) through which duplicates were removed. The first 300 abstracts were screened by 3 reviewers (HC, AS, and NR). Differences in judgment were used to refine the inclusion and exclusion criteria and were discussed until consensus was reached. This first phase of screening also served to build a common understanding of the inclusion and exclusion criteria. Screening of the remaining abstracts was performed by 1 reviewer (HC). Similarly, the first 10 full articles were screened by 2 reviewers (HC and AS), and differences in judgment were discussed until consensus was reached. If consensus was difficult to attain, a third reviewer who is a senior reviewer and an expert on usability (NR) was consulted. The remaining full papers were independently screened by one of these 3 reviewers.

To be included in this scoping review, studies had to report on user-centered usability procedures or methods of evaluation for any type of digital solution that could be relevant for older adults and that was (1) published in English; (2) a review, either systematic, scoping, or narrative review; (3) addressing and synthesizing evidence on any of the steps or methodologies used for usability assessment; and (4) addressing usability in general or for a specific digital solution that was considered relevant (this was a subjective judgment made by the authors of the review) to older adults or those caring for older adults, such as informal caregivers, family members, or health care professionals.

Studies were excluded if they (1) were grossly unrelated to the study topic (eg, chemistry field); (2) targeted children or younger age groups (eg, digital solutions for children with diabetes); (3) addressed usability for nondigital solutions (eg, buildings) or digital solutions assessed as not of interest for older adults or those caring for them (eg, moodle and eLearning solutions); and (4) addressed usability of digital solutions for caregivers of older adults, but only those studies that did not involve interaction or feedback with older persons or those caring for them were included.

Charting the Data and Collating, Summarizing, and Reporting the Results

The data extraction tool was developed using an iterative team process. The preliminary data extraction categories were derived from our research questions. The following data were extracted from each review: authors, year of publication, purpose/aim of the study, and the number of studies included in the review. Further extraction, analysis, and reporting of results were guided by the framework proposed by Ellsworth et al [29] for reporting usability evaluations, and the following operational definitions were used for this review:

1. Study evaluators, that is, the individuals who conducted the usability evaluation.
2. Participants, that is, the individuals who were asked to evaluate the usability of a product or service.

3. Tasks, that is, the activities that participants were asked to perform when evaluating the usability of a product or service.
4. Methods and techniques: methods refer to the set of techniques used to perform formative user-centered usability evaluation of a certain type at any stage of the product or service development. Usability evaluation techniques refer to a set of procedures used to perform a usability evaluation and collect data of a certain type. For this review, methods and techniques of usability evaluation were categorized and defined as presented in Table 1 (adapted from Martins et al [30]). Usability assessment usually requires the combination of more than one method, can be conducted remotely (ie, evaluators are separated in space from users) or in the presence of the participants, and can be synchronous (ie, occur at the time of the participants' interaction with the system) or asynchronous [30].
5. The test environment, that is, the environment where the evaluation of usability takes place: (1) laboratory or controlled conditions, usually a transversal assessment, or (2) in a real context, that is, the usability assessment is carried out in the same context and circumstances where the end product or service is expected to be used, which is usually a longitudinal assessment.

Details on the characteristics of each of these components of the usability assessment were extracted.

Table 1. Methods of user-centered usability evaluation.

Method and definition and technique for data collection	Definition
Test: involves observing users while they perform predefined tasks and consists of collecting mostly quantitative data; the test is centered on the interaction of the user with the technology	
Performance evaluation	Evaluated by recording elements related to the execution of a particular task (eg, execution time, success or failure, number of errors, eye-tracking, and automated usability evaluation or logfiles or web usage analysis or app-use generated data or sensor data)
Observation	Attentive visualization and systematic recording of a particular phenomenon, including people, artifacts, environments, behaviors, and interactions. Observation can be direct, when the researcher is present during the task execution, or indirect, when the task is observed through other means such as video recording
Think-aloud	Users are invited to talk about what they see, do, think, or feel as they interact with the system or service
Inquiry: provide valuable, subjective, and usually qualitative information on the users' opinions and expectations	
Focus groups	Involves a small number of people in an informal discussion
Interviews	Involves a one-to-one interaction to gather opinions, attitudes, perceptions, and experiences
Scales/questionnaires	Collects data on characteristics, thoughts, feelings, perceptions, behaviors, or attitudes, measuring either one (scale) or several (questionnaire) dimensions of usability. It is important to distinguish whether instruments were validated
Diary studies	Users record events related to their experience in the context of daily activity and later share them with the evaluators
Card sorting	It involves participants using logic while sorting content or *cards* into categories or groups that make sense to them, given the information they are provided with

Results

Overview

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for this scoping review is presented in Figure 1. A total of 3958 articles were identified from the 5 electronic databases. Of these, 1298 were eliminated because they were duplicates or did not have the author's name. The remaining 2660 records were screened based on title and abstract and 2509 were excluded because they were not reviews (66/2660, 2.48%) or were out of scope (2443/2660, 91.8%). A total of 151 full texts were read for further analysis. Of these, 115 manuscripts were excluded because they were not related to usability, 3 articles were not found, and 13 reported on the assessment of usability by experts. Therefore, 20 reviews were included in this scoping review of the reviews. Of these, 19 were systematic reviews and one was a narrative review. Table 2 presents the main characteristics of the included reviews (study, purpose, and number of included studies).

Figure 1. Flow diagram showing study identification and selection for the present review.

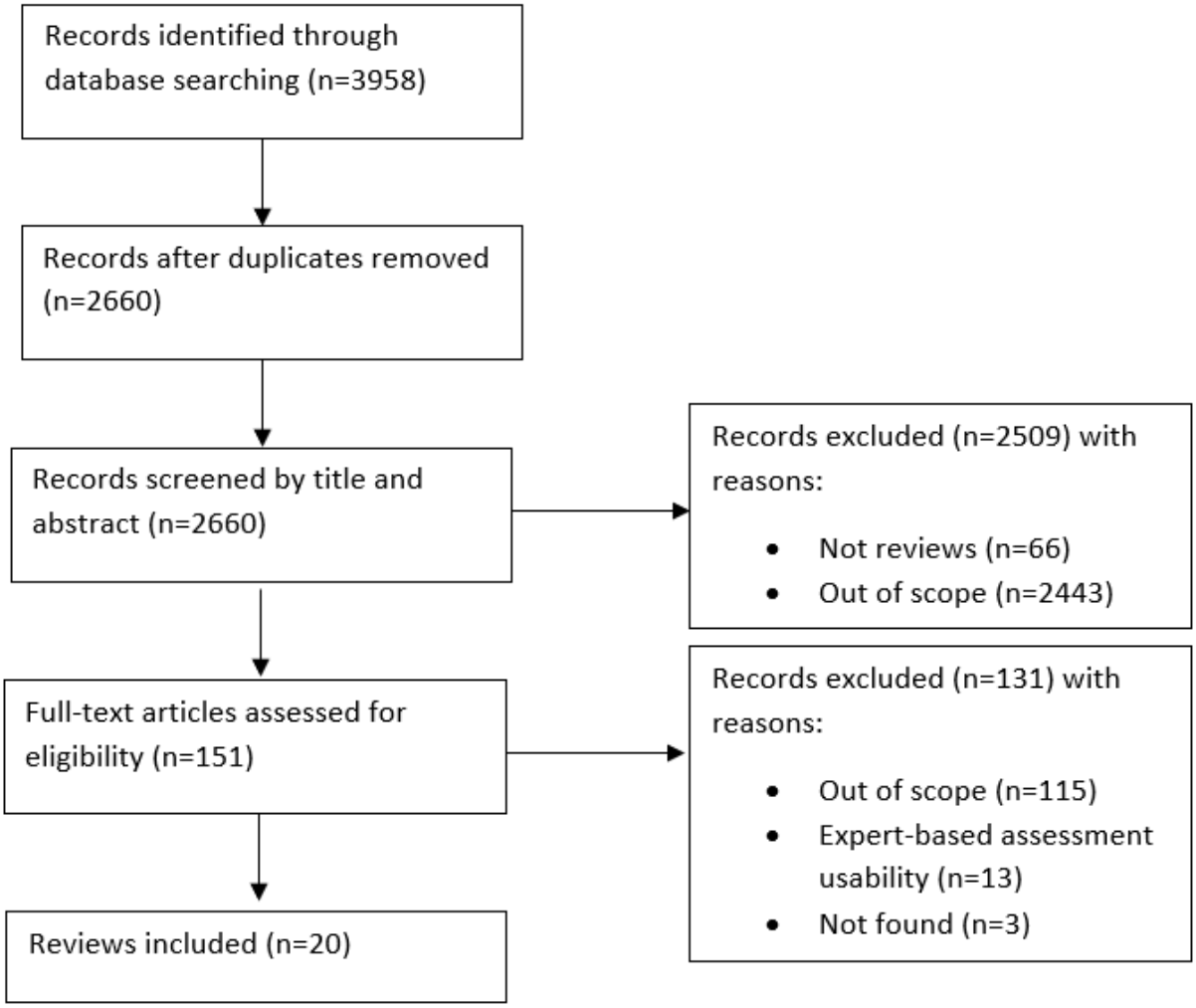


Table 2. General characteristics of included reviews.

Study	Purpose of the review	Number of studies included in the review
Ellsworth et al (2017) [29]	Review methods employed for usability testing on electronic health records	120
Allison et al (2019) [31]	Review methodologies and techniques to evaluate websites; provide a framework of the appropriate website attributes that could be applied to any future website evaluations	69
Azad-Khaneghah et al (2020) [32]	Review the rating scales used to evaluate usability and quality of mobile health applications	87
Baharuddin et al (2013) [33]	Propose a set of usability dimensions that should be considered for designing and evaluating mobile applications	Not referred
Bastien (2010) [34]	List test procedures and define and develop tools to help conduct user tests	Not referred (narrative review)
Bhutkar et al (2013) [35]	List the most commonly applied usability evaluation methods and related emerging trends	30
Cavalcanti et al (2018) [36]	Understand which methods and user assessment approaches are commonly used in motor rehabilitation studies that use augmented reality applications	32
Fernandez et al (2012) [37]	Analyze the usability evaluation methods that have proven to be the most effective in the web domain	18
Fernandez et al (2011) [38]	Analyze the usability evaluation methods that have been employed to evaluate web applications over the last 14 years	206
Fu et al (2017) [39]	Assess the usability of diabetes mobile apps developed for adults with type 2 diabetes	7
Hussain et al (2014) [40]	Review the relevant and appropriate usability dimensions and measurements for banking applications	49
Inal et al (2020) [41]	Analyze how usability is being addressed and measured in mobile health interventions for mental health problems	42
Klaassen et al (2016) [42]	Analyze if usability methods are equally employed for different end-user groups and applications	127
Lim et al (2019) [43]	Identify, study, and analyze existing usability metrics, methods, techniques, and areas in mobile augmented reality learning	72
Narasimha et al (2017) [44]	Analyzing the characteristics of usability-related studies conducted using geriatric participants and the subsequent usability challenges identified	16
Shah and Chiew (2019) [45]	Identify, analyze, and synthesize the usability features and assessment approaches of pain management mobile applications targeted at the evaluation studies	27
Simor et al (2016) [46]	Analyze usability evaluation methods used for gesture-based games, considering devices with the motion-sensing capability	10
Sousa and Lopez (2017) [47]	Identify psychometrically tested questionnaires that measure the usability of eHealth tools	35
Yen and Bakken (2012) [48]	Review and categorize health information technology usability study methods, and to provide practical guidance on health information technology usability evaluation	346
Zapata et al (2015) [49]	Review a set of selected papers that perform a usability evaluation of mobile health–related mobile apps	22

Study Evaluators

Only 4 out of the 20 (20%) [29,36,37,46] included reviews briefly mentioned any characteristic of the evaluators' profile. One of the reviews [36] reported that one of the 32 articles included mentioned that the person who performed the usability assessment was a blind evaluator. One review stated that several studies (exact numbers not provided) used graduate students as both evaluators to perform usability inspections and participants in experimental sessions (eg, think-aloud protocol, remote user

testing) [37], whereas another review [46] reported that usability evaluations were conducted by researchers. In a review by Ellsworth et al [29], 29% (35/120) of the included articles presented the description of the study evaluators responsible for designing and carrying out the usability evaluation, but the characteristics reported in primary studies were not provided.

Participants

Half of the reviews included in this scoping review did not refer to the characteristics of the participants included in the primary studies reviewed. Of the reviews, 50% (10/20) reviews that reported on any of the participants' characteristics, 4 reported mean age or age range [36,41,46,49], 4 reported the gender of participants [36,41,44,46], 8 reported the sample size [35,36,39,41,42,46,47,49], and 7 reported on other characteristics of participants by describing them as healthy participants or as having a specific clinical condition [36,37,39,41,44,46,49]. Nevertheless, 20% (4/20) reviews that reported the age of the participants also reported that not all primary studies detailed such information. Similar findings were reported for gender and sample size. No reference to sample size calculation or rationale for deciding on sample size was provided. Other characteristics of participants mentioned were being healthy, having a specific clinical condition, belonging to a specific occupational group (health care providers or students), and previous experience with mobile devices. Multimedia Appendix 1 presents a description of the information provided within the included reviews.

Tasks

Only 2 of the 20 (10%) included reviews referred to the tasks that participants were asked to perform for the usability evaluation [46,49]. Simor et al [46] conducted a usability evaluation for gesture-based games and reported that the games and, consequently, the usability evaluation of each study had different aims, target populations, interfaces, and details, but in the majority of the studies, the protocol used was presented. Zapata et al [49] performed a systematic review on mobile health apps and reported that 17 of the 22 primary studies included reported the number of tasks performed by the users. The number of tasks ranged between 1 and 25.

Methods and Techniques

Of the 20 systematic reviews included, only 3 (15%) [33,40,41] did not refer to the methods and techniques of usability used. Among the inquiry methods, the questionnaires/scales (15/20, 75%) and interviews (12/20, 60%) were most commonly reported. Among the test methods, the techniques of performance (9/20, 45%) and think-aloud were the most commonly reported (6/20, 30%; Table 3). Of the 20 reviews, 6 (30%) reported on combinations of techniques mentioning a total of 22 different combinations of 4, 3, or 2 techniques. Most combinations include at least one technique from each method, which indicates that a multimethod approach was used (Table 4). Among scales/questionnaires, which constitute the technique most often reported, the most common usability assessment scales were the System Usability Scale [29,32,41-43,46,47] and the Post-Study System Usability Questionnaire [41,42,46,47]. The other scales/questionnaires include the Questionnaire for User Interaction Satisfaction [29,42,47], the Software Usability Measurement Inventory [32,42], the Usefulness, Satisfaction, and Ease of use Questionnaire [32,41], the Computer System Usability Questionnaire [32,47], the After-Scenario Questionnaire [46,47], the Perceived Useful and Ease of Use [32], the IsoMetrics usability inventory [32], the Health Information Technology Usability Evaluation Scale [32], the user Mobile Application Rating Scale [32]; the IBM ease of use [42], and the ISO 9241–11 Questionnaire [43]. In addition, several reviews have reported the use of nonvalidated questionnaires [32,41,43,46]. One review reported that 26% of the included studies used a remote assessment of usability, where participants are in an uncontrolled environment [31].

Table 3. Detailed techniques used for usability evaluation.

Study	Test			Inquiry				
	Performance evaluation (n=9)	Observation (n=3)	Think-aloud (n=6)	Focus group (n=3)	Interview (n=12)	Scales or questionnaires (n=15)	Diary studies (n=1)	Card sorting (n=1)
Allison et al (2019) [31]	✓[a]	—[b]	—	—	—	✓	—	—
Azad-Khaneghah et al (2020) [32]	—	—	—	—	—	✓	—	—
Bastien (2010) [34]	—	—	—	—	✓	—	✓	—
Bhutkar et al (2013) [35]	✓	—	✓	—	✓	—	—	—
Cavalcanti et al (2018) [36]	✓	—	✓	—	—	✓	—	—
Ellsworth et al (2017) [29]	—	—	—	✓	✓	✓	—	✓
Fernandez et al (2012) [37]	✓	—	✓	—	✓	✓	—	—
Fernandez et al (2011) [38]	✓	—	✓	✓	✓	✓	—	—
Fu et al (2017) [39]	✓	—	—	—	—	✓	—	—
Klaassen et al (2016) [42]	✓	✓	—	—	✓	✓	—	—
Lim et al (2019) [43]	✓	—	—	—	✓	✓	—	—
Narasimha et al (2017) [44]	—	—	—	—	✓	✓	—	—
Shah and Chiew (2019) [45]	—	✓	—	—	✓	✓	—	—
Simor et al (2016) [46]	—	—	—	—	✓	✓	—	—
Sousa and Lopez (2017) [47]	—	—	—	—	—	✓	—	—
Yen and Bakken (2012) [48]	—	✓	✓	✓	✓	✓	—	—
Zapata et al (2015) [49]	✓	—	✓	—	✓	✓	—	—

[a]Reported in the review.

[b]Not reported.

Table 4. Detailed description of the combination of techniques used for usability assessment.

Techniques	Study						Multimethod
	Cavalcanti et al (2018) [36]	Fu et al (2017) [39]	Inal et al (2020) [41]	Shah & Chiew (2019) [45]	Simor et al (2016) [46]	Zapata et al (2015) [49]	
Observation + performance evaluation + think-aloud + scale/questionnaire	✓[a]	N/A[b]	N/A	N/A	N/A	N/A	✓
Observation + performance evaluation + scale/questionnaire + interview	✓	N/A	N/A	N/A	N/A	N/A	✓
Observation + scale/questionnaire+ interview + diary studies	N/A	N/A	✓	N/A	N/A	N/A	✓
Performance evaluation + think-aloud + scale/questionnaire + interview	N/A	N/A	✓	N/A	N/A	N/A	✓
Observation + performance evaluation + think-aloud + interview	N/A	N/A	✓	N/A	N/A	N/A	✓
Performance evaluation + scale/questionnaire + interview	✓	N/A	✓	✓	N/A	N/A	✓
Performance evaluation + scale/questionnaire + focus group	N/A	N/A	✓	N/A	N/A	N/A	✓
Performance evaluation + scale/questionnaire + observation	✓	N/A	✓	N/A	N/A	N/A	✓
Performance evaluation + observation	N/A	N/A	✓	N/A	✓	N/A	N/A
Think-aloud + scale/questionnaire + interview	N/A	N/A	✓	✓	N/A	N/A	✓
Think-aloud + scale/questionnaire + interview	N/A	✓	N/A	N/A	N/A	N/A	✓
Scale/questionnaire + interview + focus group	N/A	N/A	✓	✓	N/A	N/A	N/A
Observation + scale/questionnaire + interview	✓	N/A	✓	✓	N/A	N/A	✓
Observation + scale/questionnaire	✓	✓	✓	N/A	N/A	N/A	✓
Observation + interview	N/A	N/A	N/A	✓	N/A	N/A	✓
Performance evaluation + observation	✓	N/A	N/A	N/A	N/A	N/A	N/A
Performance evaluation + scale/questionnaire	✓	N/A	✓	N/A	N/A	✓	✓
Think-aloud + scale/questionnaire	N/A	N/A	✓	N/A	N/A	N/A	✓
Think-aloud +interview	N/A	N/A	N/A	✓	N/A	N/A	✓
Scale/questionnaire + interview	✓	N/A	✓	✓	N/A	✓	N/A
Scale/questionnaire + diary studies	N/A	N/A	✓	N/A	N/A	N/A	N/A
Interview + focus group	N/A	N/A	✓	N/A	N/A	N/A	N/A

[a]Reported in the review.

[b]N/A: not applicable.

Test Environment

Of the 20 reviews, 2 (10%) reported on the environment where the usability assessment of the included studies took place. In a review by Bhutkar et al [35], of the 17 studies that reported on the test environment, 8 were conducted in hospitals, 5 in intensive care units, and 4 in laboratories. In addition, 31 of the 42 studies reviewed by Inal et al [41], which focused on mobile health interventions for mental health problems, reported having conducted their usability testing in the natural environment of the participants with the technology deployed in the everyday environment of the intended users or their representatives. In addition, the review of Ellsworth et al [29] did not provide data on the test environment; however, the test environment was an inclusion criterion, as they stated that they have included studies that tested the usability of the hospital and clinic electronic health records in the inpatient, outpatient, emergency department, or operating room settings.

Discussion

Principal Findings

This scoping review of reviews aims to synthesize the procedures used or reported for the different steps of the process of conducting a user-centered usability assessment of digital solutions relevant for older adults, to identify gaps in the literature, and to identify the best practices for each of the different steps. The results suggest that the characteristics of study evaluators and participants and task procedures are only briefly reported, and no agreement seems to exist on what should be reported. The methods and techniques used for the assessment of user-centered usability are the topics most commonly and comprehensively reported in the reviews, whereas the test environment is seldom and poorly characterized. Despite our aim of searching for reviews reporting on digital solutions relevant for older adults, only one of the included reviews specifically targeted older adults. This suggests that studies using older adults are scarce and that the findings of this scoping review also apply to usability studies with adults.

Our findings are in line with the review of Ellsworth et al [29], who reported that several of the included studies described the participants, but not the individual who conducted the usability assessment (study evaluator). The level of expertise and domain experience, whether the study evaluator is external to the team developing the product or service being assessed or, on the contrary, is part of the team and potentially has a conflict of interest when assessing usability, are examples of aspects that have the potential to influence the results of the usability assessment. Therefore, these should be reported by the authors. Most of the techniques are complex procedures of usability assessment; some of these depend on the interaction between the participant and the study evaluator and, therefore, require experience and knowledge to be assessed effectively.

The characteristics of the study participants most commonly reported across reviews were age and sex. However, these seem insufficient for the reader to make a judgment regarding the degree of similarity between the sample and the target end users. Educational or digital literacy levels are likely to influence how the participant perceives the usability of the system. For example, different subgroups of older adults may perceive the usability of the same system differently [46]. Therefore, a detailed characterization of physical, emotional, cognitive, and digital skills is needed for an appropriate interpretation of the results of the usability evaluation in certain subgroups of older adults. Furthermore, a detailed characterization of health conditions might also be relevant [46]. These aspects will also inform whether the sample used is representative of the end users. The use of nonrepresentative users and, therefore, the failure to consider their needs and preferences may result in products with low usability [36]. In general, the sample sizes are small, and no rationale for the size of the sample is provided. The appropriate sample size for usability studies is a matter of debate, with some authors arguing that 4 or 5 participants are enough to identify approximately 80%-85% of usability problems [50-52], whereas others report that with these numbers of participants only 35% of usability problems are determined [53]. The type of interfaces, the tasks performed by the participants, the context of use, and the state of technology development may explain the differences between studies [34]. Furthermore, it is worth noting the definition of usability as the measure by which a product can be used by specific users to achieve specific goals with effectiveness, efficiency, and satisfaction [15]. Conceivably, small sample sizes may be enough to detect usability problems but may be insufficient to have a broader view of usability more in line with the present definition.

Only 2 reviews reported on the tasks that participants were asked to perform to assess the usability of the product or service [46,49], and both concluded that, in general, studies reported on the protocol of the tasks used. Tasks vary depending on several factors, such as study aims, target population, interfaces, methods, and techniques used for usability assessment [46]. Nevertheless, the definition or selection of tasks that participants should perform should mirror the future use of the product or service [34,40]. No principles were found to guide the selection of tasks. For example, should there be a minimum set of tasks to be performed, should tasks require single or multiple steps, or should there be a minimum amount of time that each participant needs to spend using the product or service are illustrative examples of issues that are not clear.

The methods and techniques used for the assessment of usability have been consistently reported, and most reviews have found that a combination of methods and/or techniques are usually performed, in line with recommendations [19]. Different methods and techniques have different strengths and limitations [46] and, therefore, their combination is more likely to provide a comprehensive view of usability problems [19]. For example, scales and questionnaires are easy to use and useful for gathering self-reported data about the user's perception but might have limited value informing on which aspects of the system need to be targeted for improvement [29,54]. Scales and questionnaires should be valid, but a few reviews have reported the use of scales and questionnaires that are unlikely to have been validated. Although there might be reasons to develop or adapt a scale/questionnaire, this process must be followed by evidence of its validity [41]. Interviews and observations are recommended when the number of participants is small because

both generate high amounts of data that are time-consuming to analyze. Nevertheless, interviews can be useful to understand the reasoning of the user when facing a problem, and observation gives an insight into the moment when a problem occurs [46]. It is argued that think-aloud protocols may result in the loss of focus on the tasks being performed, whereas user performance is an easy assessment, particularly in cases where the system automatically records the performance indicators, but might provide limited information if used alone [46]. The most frequent multimethod combination described in the literature is the test and inquiry method combination; however, we found no information in the included reviews regarding which combination of techniques is the most sensitive and whether this could vary depending on the development stage of the product or service being evaluated. Furthermore, the combination of techniques should allow for the assessment of effectiveness, efficiency, and satisfaction, as these are all part of usability.

Only 2 reviews reported on the test environment, but both referred that most included studies reported usability testing to have been conducted in the real context. Nevertheless, we found no indication of how long the usability assessment should be conducted, that is, how long the participants should be allowed to use the product or service before assessing it, and whether conducting the usability assessment in a real context means that the product or service was used in the circumstances that it is expected to be used.

Recommendations and Future Research

The conducting of rigorous experiments on user-centered usability is likely to result in increased sensitivity for these experiments, that is, an increased ability to detect usability issues. Developing a consensus framework is likely to improve the quality of studies on usability evaluation and respective reporting, improve comparability of usability results across studies, provide digital solutions helping consumers and producers to identify the best products, improve the efficiency of the process of usability evaluation and facilitate further research on the impact of usability on other outcomes, such health-related outcomes. Textbox 1 presents a list of parameters that we believe should be considered when planning and reporting user-centered usability studies. These parameters provide guidance while also being flexible to accommodate study differences regarding aspects such as study participants or the digital solution being assessed. At present, we are working on a Delphi-study aiming to establish an international consensus on user-centered usability evaluation procedures.

Textbox 1. A proposed guide of aspects to consider when designing and reporting a user-centered usability evaluation study.

Study evaluator:

- Provide a rationale for sample size
- Experience with usability evaluation with users (if none, plan training)
- Establish clear inclusion and exclusion criteria (age, gender, educational level, and academic background)
- Clarify whether internal or external to product development

Participants:

- Provide a rationale for sample size
- Define clear inclusion and exclusion criteria
- Define sampling methods (probability/nonprobability) and setting of recruitment

Methods and techniques:

- Provide a rationale for the combination of methods and techniques
- Define equipment needed
- Select valid and reliable instruments of assessment

Task:

- Define the number
- Provide a detailed description of tasks
- Develop a participant script

Test environment/equipment:

- Identify and justify the choice (lab test or field test or both; remote test or face to face)
- Identify facilities and material needed
- Ensure the existence of an observation room and recording room
- Ensure the proper functioning of all equipment necessary for the test evaluation

Limitations of This Scoping Review

Some limitations are directly related to the typology of this review, such as the absence of assessment of the quality of the included reviews and the quantitative summary of findings [55]. Usability is also a topic on which a large number of publications are published as conference proceedings, and such publications were not specifically searched (selection bias). Nevertheless, it is likely that by including mostly reviews published in journals that these are more comprehensive, as conference proceedings tend to have lower word counts for included papers. Abstracts and full-text screening were performed first by 3 and 2 authors, respectively, and after a common understanding was built, only 1 reviewer screened the remaining abstracts and full papers. Although we believe that this did not have a major impact on the results, having only 1 person screening for inclusion might have increased the possibility of error and of not including a potentially relevant study. The judgment made to decide whether a manuscript was on a product or technology that could be of use for older adults was a subjective judgment made by the authors and could have biased the results toward the field of health. Finally, no cross-checking of the primary studies included in each review was made and, therefore, the same primary studies could have been included in more than one review.

In summary, we found a lack of a detailed description of several steps of the process of assessing the usability of digital solutions and no evidence on good practices. These findings suggest the need for a consensus framework on the assessment of usability that informs researchers and allows standardization of procedures. Furthermore, it highlights the need to investigate whether different techniques of assessing usability are more sensitive than others to detect usability issues.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary details of participant profile and sample size (sometimes percentages do not add up to 100%, as only partial information was provided in the review).
[DOCX File , 31 KB - humanfactors_v8i1e22774_app1.docx]

References

1. Beschorner B, Woodward L. Engaging teachers in a digital learner-centered approach to support understanding foundational literacy. In: Karchmer-Klein R, Pytash KE, editors. Effective Practices in Online Teacher Preparation for Literacy Educators. Pennsylvania: IGI Global; 2020:284-306.
2. Wagner F, Basran J, Bello-Haas VD. A review of monitoring technology for use with older adults. J Geriatr Phys Ther 2012;35(1):28-34. [doi: 10.1519/JPT.0b013e318224aa23] [Medline: 22189952]
3. Simões P, Silva AG, Amaral J, Queirós A, Rocha NP, Rodrigues M. Features, behavioral change techniques, and quality of the most popular mobile apps to measure physical activity: systematic search in app stores. JMIR Mhealth Uhealth 2018 Oct 26;6(10):e11281 [FREE Full text] [doi: 10.2196/11281] [Medline: 30368438]
4. Silva AG, Simões P, Queirós A, Rodrigues M, Rocha NP. Mobile apps to quantify aspects of physical activity: a systematic review on its reliability and validity. J Med Syst 2020 Jan 8;44(2):51. [doi: 10.1007/s10916-019-1506-z] [Medline: 31915935]
5. Gilson A, Dodds D, Kaur A, Potteiger M, Ii JH. Using computer tablets to improve moods for older adults with dementia and interactions with their caregivers: pilot intervention study. JMIR Form Res 2019 Sep 3;3(3):e14530 [FREE Full text] [doi: 10.2196/14530] [Medline: 31482847]
6. Leone C, Lim JS, Stern A, Charles J, Black S, Baecker R. Communication technology adoption among older adult veterans: the interplay of social and cognitive factors. Aging Ment Health 2018 Dec;22(12):1666-1677. [doi: 10.1080/13607863.2017.1381946] [Medline: 29019416]
7. Rogers WA, Mitzner TL. Envisioning the future for older adults: autonomy, health, well-being, and social connectedness with technology support. Futures 2017 Mar;87:133-139 [FREE Full text] [doi: 10.1016/j.futures.2016.07.002] [Medline: 28458395]
8. Fortin M, Hudon C, Haggerty J, van den Akker M, Almirall J. Prevalence estimates of multimorbidity: a comparative study of two sources. BMC Health Serv Res 2010 May 6;10:111 [FREE Full text] [doi: 10.1186/1472-6963-10-111] [Medline: 20459621]
9. Fortin M, Bravo G, Hudon C, Vanasse A, Lapointe L. Prevalence of multimorbidity among adults seen in family practice. Ann Fam Med 2005;3(3):223-228 [FREE Full text] [doi: 10.1370/afm.272] [Medline: 15928225]
10. de la Torre-Díez I, López-Coronado M, Vaca C, Aguado JS, de Castro C. Cost-utility and cost-effectiveness studies of telemedicine, electronic, and mobile health systems in the literature: a systematic review. Telemed J E Health 2015 Feb;21(2):81-85 [FREE Full text] [doi: 10.1089/tmj.2014.0053] [Medline: 25474190]
11. Hajek A, Jens-Oliver B, Kai-Uwe S, Matschinger H, Brenner H, Holleczek B, et al. Frailty and healthcare costs-longitudinal results of a prospective cohort study. Age Ageing 2018 Mar 1;47(2):233-241. [doi: 10.1093/ageing/afx157] [Medline: 29036424]

12. Cheung JT, Yu R, Wu Z, Wong SY, Woo J. Geriatric syndromes, multimorbidity, and disability overlap and increase healthcare use among older Chinese. BMC Geriatr 2018 Jun 25;18(1):147 [FREE Full text] [doi: 10.1186/s12877-018-0840-1] [Medline: 29940868]
13. Herndon JH, Hwang R, Bozic KH. Healthcare technology and technology assessment. Eur Spine J 2007 Aug;16(8):1293-1302 [FREE Full text] [doi: 10.1007/s00586-007-0369-z] [Medline: 17426985]
14. Pyla PS, Hartson R. Agile UX Design for a Quality User Experience. Massachusetts: Morgan Kaufmann Publishers; 2019.
15. ISO 9241-11:2018(en) Ergonomics of Human-system Interaction — Part 11: Usability: Definitions and Concepts. ISO. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2020-10-01]
16. Berntsen NO, Dybkjær L. Multimodal Usability. London, UK: Springer; 2010.
17. Nunes IL. Ergonomics and usability – key factors in knowledge society. Enterp Work Innov Stud 2006;2:87-94.
18. Martins A, Queirós A, Silva AG, Rocha NP. Usability evaluation methods: a systematic review. In: Saeed S, Bajwa IS, Mahmood Z, editors. Human Factors in Software Development and Design. Pennsylvania: IGI Global; 2014.
19. Morrissey K. A review of 'universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions'. Visitor Studies 2014 Sep 25;17(2):222-224. [doi: 10.1080/10645578.2014.945354]
20. Dix A, Finlay G, Abowd G, Beale R. Human-Computer Interaction. New Jersey, US: Prentice Hall; 2004.
21. da Costa RP, Canedo ED, de Sousa RT, de Oliveira RA, Villalba LJ. Set of usability heuristics for quality assessment of mobile applications on smartphones. IEEE Access 2019;7:116145-116161. [doi: 10.1109/access.2019.2910778]
22. Dix A. In: Finlay G, Abowd G, Beale R, editors. Human-Computer Interaction. New Jersey, US: Prentice Hall; 2004.
23. Bernsen N. In: Dybkjær L, editor. Multimodal Usability. London, UK: Springer; 2010.
24. Martins AI, Queirós A, Rocha NP, Santos BS. Avaliação de usabilidade: uma revisão sistemática da literatura. Iber J Inf Syst Technol 2013 Jun 1;11:31-44. [doi: 10.4304/risti.11.31-44]
25. Queirós A, Silva A, Alvarelhão J, Rocha NP, Teixeira A. Usability, accessibility and ambient-assisted living: a systematic literature review. Univ Access Inf Soc 2013 Oct 5;14(1):57-66. [doi: 10.1007/s10209-013-0328-x]
26. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
27. 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]
28. Ambient Assisted Living. AAL Programme. 2007. URL: http://www.aal-europe.eu/ [accessed 2020-09-30]
29. Ellsworth MA, Dziadzko M, O'Horo JC, Farrell AM, Zhang J, Herasevich V. An appraisal of published usability evaluations of electronic health records via systematic review. J Am Med Inform Assoc 2017 Jan;24(1):218-226 [FREE Full text] [doi: 10.1093/jamia/ocw046] [Medline: 27107451]
30. Martins AI. In: Queirós A, Silva AG, Rocha NP, editors. Usability Evaluation Methods: A Systematic Review, Human Factors. Pennsylvania: IGI Global; 2014.
31. Allison R, Hayes C, McNulty CA, Young V. A comprehensive framework to evaluate websites: literature review and development of goodweb. JMIR Form Res 2019 Oct 24;3(4):e14372 [FREE Full text] [doi: 10.2196/14372] [Medline: 31651406]
32. Azad-Khaneghah P, Neubauer N, Cruz AM, Liu L. Mobile health app usability and quality rating scales: a systematic review. Disabil Rehabil Assist Technol 2020 Jan 8:1-10. [doi: 10.1080/17483107.2019.1701103] [Medline: 31910687]
33. Baharuddin R, Singh D, Razali R. Usability dimensions for mobile applications-a review. Res J Appl Sci Eng Technol 2013 Feb 21;11(9):2225-2231. [doi: 10.19026/rjaset.5.4776]
34. Bastien JC. Usability testing: a review of some methodological and technical aspects of the method. Int J Med Inform 2010 Apr;79(4):e18-e23. [doi: 10.1016/j.ijmedinf.2008.12.004] [Medline: 19345139]
35. Bhutkar G, Konkani A, Katre D, Ray GG. A review: healthcare usability evaluation methods. Biomed Instrum Technol 2013;Suppl:45-53. [doi: 10.2345/0899-8205-47.s2.45] [Medline: 24111771]
36. Cavalcanti VC, Santana MI, Gama AE, Correia WF. Usability assessments for augmented reality motor rehabilitation solutions: a systematic review. Int J Comput Games Technol 2018 Nov 1;2018:1-18. [doi: 10.1155/2018/5387896]
37. Fernandez A, Abrahão S, Insfran E. A Systematic Review on the Effectiveness of Web Usability Evaluation Methods. In: 16th International Conference on Evaluation & Assessment in Software Engineering. 2012 Presented at: EASE'12; June 1-7, 2012; Ciudad Real. [doi: 10.1049/ic.2012.0007]
38. Fernandez A, Insfran E, Abrahão S. Usability evaluation methods for the web: a systematic mapping study. Inf Softw Technol 2011 Aug;53(8):789-817. [doi: 10.1016/j.infsof.2011.02.007]
39. Fu H, McMahon SK, Gross CR, Adam TJ, Wyman JF. Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: a systematic review. Diabetes Res Clin Pract 2017 Sep;131:70-81. [doi: 10.1016/j.diabres.2017.06.016] [Medline: 28692830]
40. Hussain A, Abubakar HI, Hashim B. Evaluating Mobile Banking Application: Usability Dimensions and Measurements. In: 6th International Conference on Information Technology and Multimedia. 2014 Presented at: CITM'14; May 20-23, 2014; Putrajaya. [doi: 10.1109/icimu.2014.7066618]
41. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. J Med Internet Res 2020 Jan 6;22(1):e15337 [FREE Full text] [doi: 10.2196/15337] [Medline: 31904579]

42. Klaassen B, van Beijnum BJ, Hermens HJ. Usability in telemedicine systems-a literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
43. Lim KC, Selamat A, Alias RA, Krejcar O, Fujita H. Usability measures in mobile-based augmented reality learning applications: a systematic review. Appl Sci 2019 Jul 5;9(13):2718. [doi: 10.3390/app9132718]
44. Narasimha S, Madathil KC, Agnisarman S, Rogers H, Welch B, Ashok A, et al. Designing telemedicine systems for geriatric patients: a review of the usability studies. Telemed J E Health 2017 Jun;23(6):459-472. [doi: 10.1089/tmj.2016.0178] [Medline: 27875667]
45. Shah U, Chiew T. A systematic literature review of the design approach and usability evaluation of the pain management mobile applications. Symmetry 2019 Mar 19;11(3):400. [doi: 10.3390/sym11030400]
46. Simor FW, Brum MR, Schmidt JD, Rieder R, de Marchi AC. Usability evaluation methods for gesture-based games: a systematic review. JMIR Serious Games 2016 Oct 4;4(2):e17 [FREE Full text] [doi: 10.2196/games.5860] [Medline: 27702737]
47. Sousa V, Lopez KD. Towards usable E-health. A systematic review of usability questionnaires. Appl Clin Inform 2017 May 10;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]
48. Po-Yin Y, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
49. 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]
50. Nielsen J, Landauer TK. A Mathematical Model of the Finding of Usability Problems. In: Conference on Human Factors in Computing Systems. 1993 Presented at: CHI'93; July 6-9, 1993; Amsterdam. [doi: 10.1145/169059.169166]
51. Virzi RA. Streamlining the design process: running fewer subjects. Proc Hum Factors Soc Annu Meet 2016 Aug 6;34(4):291-294. [doi: 10.1177/154193129003400411]
52. Virzi RA. Refining the test phase of usability evaluation: how many subjects is enough? Hum Factors 2016 Nov 23;34(4):457-468. [doi: 10.1177/001872089203400407]
53. Spool J, Schroeder W. Testing Web Sites: Five Users Is Nowhere Near Enough. In: Human Factors in Computing Systems. 2001 Presented at: HFCS'01; August 5-7, 2001; Seattle. [doi: 10.1145/634067.634236]
54. Martins AI, Queirós A, Silva AG, Rocha NP. ICF Based Usability Scale: Evaluating Usability According to the Evaluators' Perspective About the Users' Performance. In: 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion. 2016 Presented at: SDTE'16; March 2-3, 2016; Vila Real. [doi: 10.1145/3019943.3019997]
55. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]

Edited by G Eysenbach; submitted 23.07.20; peer-reviewed by S Olsen, T Risling, A Jaiswal; comments to author 16.09.20; revised version received 29.09.20; accepted 29.10.20; published 13.01.21.

Please cite as:
Silva AG, Caravau H, Martins A, Almeida AMP, Silva T, Ribeiro Ó, Santinha G, Rocha NP
Procedures of User-Centered Usability Assessment for Digital Solutions: Scoping Review of Reviews Reporting on Digital Solutions Relevant for Older Adults
JMIR Hum Factors 2021;8(1):e22774
URL: http://humanfactors.jmir.org/2021/1/e22774/
doi:10.2196/22774
PMID:33439128

Review

The Use of Telehealth Technology to Support Health Coaching for Older Adults: Literature Review

Carl Markert[1], BS, MBA; Farzan Sasangohar[1,2], BA, BCS, MASc, SM, PhD; Bobak J Mortazavi[3], BA, BS, MS, PhD; Sherecce Fields[4], BA, BS, MA, PhD

[1]Department of Industrial and Systems Engineering, Texas A&M University, College Station, TX, United States

[2]Center for Outcome Research, Houston Methodist, Houston, TX, United States

[3]Department of Computer Science and Engineering, Texas A&M University, College Station, TX, United States

[4]Department of Psychological and Brain Sciences, Texas A&M University, College Station, TX, United States

Corresponding Author:
Farzan Sasangohar, BA, BCS, MASc, SM, PhD
Department of Industrial and Systems Engineering
Texas A&M University
3131 TAMU
College Station, TX, 77843
United States
Phone: 1 9794582337
Email: sasangohar@tamu.edu

Abstract

Background: Health coaching is an intervention process for driving behavior change through goal-setting, education, encouragement, and feedback on health-related behaviors. Telehealth systems that include health coaching and remote monitoring are making inroads in managing chronic conditions and may be especially suited for older populations.

Objective: This literature review aimed to investigate the current status of health coaching interventions incorporating telehealth technology and the associated effectiveness of this intervention to deliver health care with an emphasis on older adults (aged 65 and older).

Methods: A literature review was conducted to identify the research conducted on health coaching combined with remote monitoring for delivering health care to older adults. The Ovid MEDLINE and CINAHL databases were queried using a combination of relevant search terms (including middle aged, aged, older adult, elderly, health coaching, and wellness coaching). The search retrieved 196 papers published from January 2010 to September 2019 in English. Following a systematic review process, the titles and abstracts of the papers retrieved were screened for applicability to health coaching for older adults to define a subset for further review. Papers were excluded if the studied population did not include older adults. The full text of the 42 papers in this subset was then reviewed, and 13 papers related to health coaching combined with remote monitoring for older adults were included in this review.

Results: Of the 13 studies reviewed, 10 found coaching supported by telehealth technology to provide effective outcomes. Effectiveness outcomes assessed in the studies included hospital admissions/re-admissions, mortality, hemoglobin A_{1c} (HbA_{1c}) level, body weight, blood pressure, physical activity level, fatigue, quality of life, and user acceptance of the coaching program and technology.

Conclusions: Telehealth systems that include health coaching have been implemented in older populations as a viable intervention method for managing chronic conditions with mixed results. Health coaching combined with telehealth may be an effective solution for providing health care to older adults. However, health coaching is predominantly performed by human coaches with limited use of technology to augment or replace the human coach. The opportunity exists to expand health coaching to include automated coaching.

(JMIR Hum Factors 2021;8(1):e23796) doi:10.2196/23796

KEYWORDS

telemedicine; remote sensing technology; health coaching; decision support systems; clinical; older adults

Introduction

Overview of Chronic Diseases

Chronic diseases are health-related conditions that require ongoing medical attention or limit one's daily activities [1]. These conditions are common among older adults and were the leading causes of death among older adults (aged 65 and older) in the United States in 2017 [2]. Chronic disease management within the world's aging population is creating a burden on the health care industry [3]. For example, the average medical expenditures in the United States within this older population were 2.6 times the national average and accounted for over one-third of medical spending in 2010 [4]. A subsequent survey by the Kaiser Family Foundation found that older adults (age 55 and over) in the United States accounted for 56% of all health care spending in 2016 but made up only 29% of the population [5].

The Census Bureau projects that the US population aged 65 or older will grow from 49 million in 2016 to 95 million by 2060 [6]. Ninety percent of these older adults prefer to *age in place*, or remain in their homes as they grow older [7] which could also mitigate health care costs for this population compared to the cost of assisted living communities. Aging in place allows them to better maintain contact with friends and family, but this preference presents a challenge for determining what health-related technology is needed to help meet this desire [8]. Telehealth may be one way to effectively manage chronic diseases among older adults while also enabling them to live at home, especially with a number of opportunities available to assist aging in place through advancements in smart sensing technology [9]. Furthermore, the COVID-19 pandemic has also shown the necessity of understanding the efficacy of telehealth systems, as these systems may be the only mode of non-emergency health care delivery for vulnerable populations in a pandemic situation [10]. However, despite the increased access to telehealth technologies, implementation strategies that do not address self-management of one's health care have led to disappointing findings, such as the failure to reduce re-admissions in individuals with heart failure [11,12].

While telehealth has enabled virtual visits with health care professionals, the self-management capabilities of telehealth require special attention to patient engagement and behavior change methods to improve active participation. Health coaching has gained widespread use in the past few years. Two recent systematic reviews found health coaching to be somewhat effective for adults with chronic conditions [13,14]. Kivelä et al [13] found health coaching to be effective for the patient's physiological, behavioral, and psychological status, specifically, improvements in weight management, physical activity, physical health, and mental health. Oliveira et al [14] found health coaching to be effective in increasing the level of physical activity in older adults but found no significant improvement in quality of life, mobility, or mood. Neither of these studies evaluated health coaching combined with remote monitoring. The goal of our review was to investigate the current status of health coaching interventions that incorporate telehealth remote monitoring technology and the associated effectiveness of this intervention with an emphasis on older adults.

Background

Telehealth is an all-encompassing term for clinical and nonclinical remote health care services and is defined by the Center for Connected Health Policy as "a collection of means or methods for enhancing health care, public health and health education delivery and support using telecommunications technologies" [15]. For the purpose of this literature review, telehealth includes telemedicine, remote patient monitoring (RPM), remote activity monitoring (RAM), decision support systems (DSSs), and health coaching systems.

Telemedicine is the use of telecommunication technology to allow health care workers to provide clinical services (eg, medical therapy) to patients remotely [16]. Telemedicine is useful for providing clinical services to patients in sparsely populated areas or places remotely located from a health care facility [17].

RPM is the use of electronic devices and telecommunication technology to monitor and transmit patient physiological or metabolic parameters to a digital database that can be accessed by authorized users [18]. RPM usually involves Bluetooth-enabled or internet-connected devices that automatically transmit monitored parameters. RPM can also include electronic wellness questionnaires that elicit information concerning the patient's well-being and health status.

RAM is the use of electronic devices to provide remote monitoring of a person's mobility or activities of daily living (ADLs) [19]. ADLs can be remotely monitored using motion detection devices installed in a person's residence or a wearable device, such as a smart watch, that detects, records, and transmits movement activity. Another form of ADL monitoring is medication adherence monitored remotely via automated pillboxes. Automated pillboxes are used to organize medications, provide reminders to take medications, and provide information to clinicians via telehealth regarding medication use [20].

DSSs are electronic (computerized) systems which evaluate data collected via remote monitoring and transform the data into useful information regarding the patient's health and wellness [21]. The DSS makes clinical or behavioral recommendations based on an evaluation of the monitored data. An example of a recommendation is a reminder to the patient to take his/her medication if an automated pillbox senses the person has not taken their medication that day. If the medication is still not taken after some delay, the DSS can notify the health care providers or health coaching system. The DSS can also initiate an emergency notification to 911 if certain threshold values of monitored parameters are exceeded.

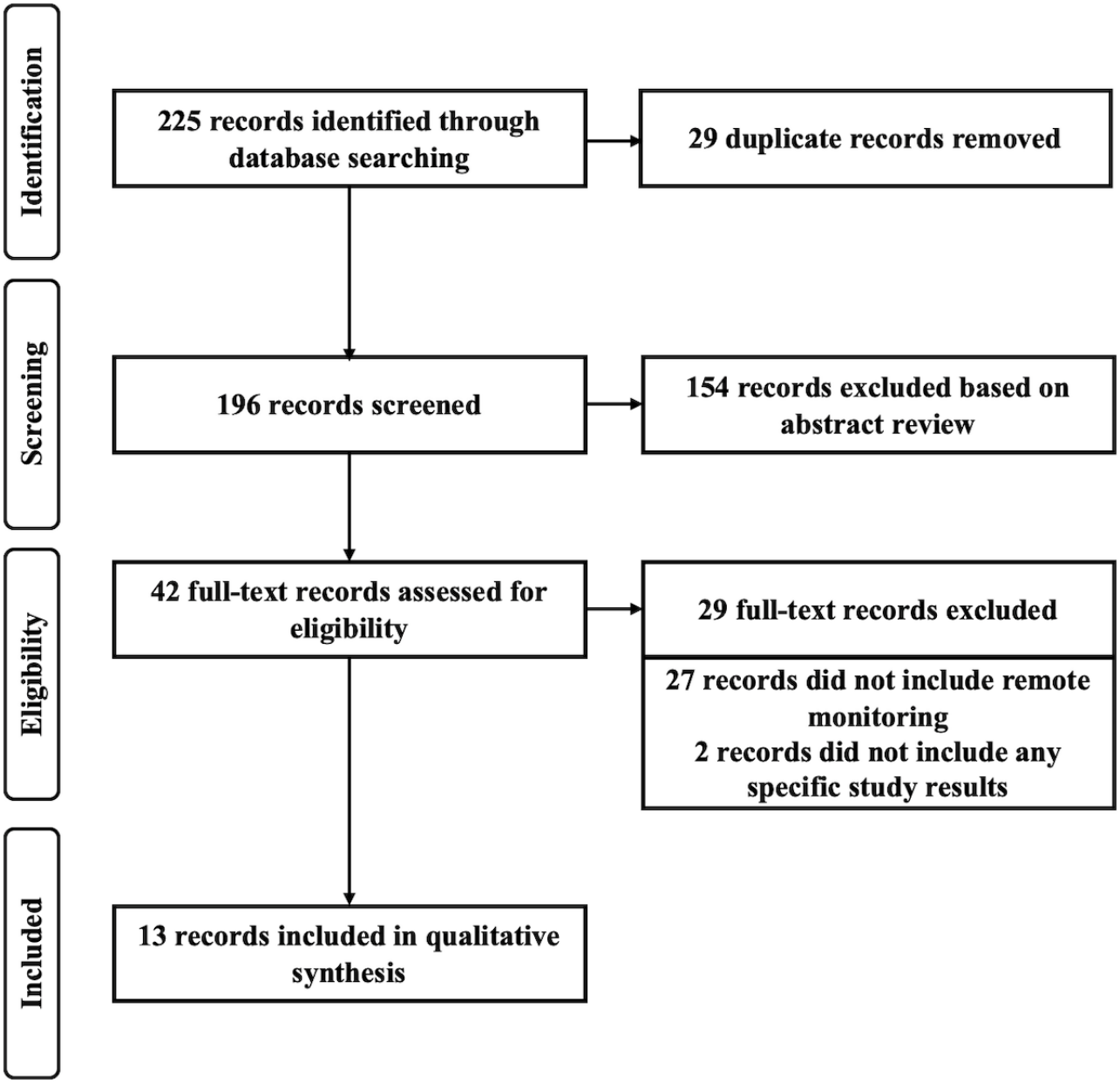
Health coaching systems are defined as "patient-centered processes that are based upon behavior change theory" and include goal setting, education, encouragement, and feedback on health-related behaviors [14]. Disease management, by contrast, focuses on the specific disease(s) instead of the patient's behavior [22]. Health coaching programs provide health-related information, recommendations, or encouragement to the patient on a routine or as-needed basis to help drive

behavior changes [21]. Forms of health coaching include encouragement, feedback, health care suggestions, periodic health tips, or short educational presentations based on an analysis of the patient's health status and monitored data. An example of a coaching message is sleep management advice if the patient is not sleeping well. The health coaching system can be manual (human health coach only), partially automated, or fully automated using artificial intelligence and machine learning to generate health coaching messages to the patient.

Methods

A literature review was chosen for this study to identify the research conducted on the current state and effectiveness of health coaching combined with remote monitoring (RPM or RAM) and any knowledge gaps that warrant further research. This review was specifically focused on health coaching combined with telehealth to deliver health care with an emphasis on older adults. The Ovid MEDLINE and CINAHL databases were queried to first retrieve papers related to health or wellness coaching for populations that included older adults and to then narrow the results to those studies that included some form of remote monitoring. Given the rapid pace with which telehealth is advancing, results from 2010 or later were chosen for this search to focus on relatively current research. The full electronic search strategy was [(MH "Middle Aged") OR (MH "Aged+") OR AB (older adult* or elder* or aged) OR TI (older adult* or elder* or aged)] AND [AB ((health or wellness) n1 coaching) OR TI ((health or wellness) n1 coaching)]. The search criteria included articles published from January 2010 to September 5, 2019 (date of search) in English. Keywords included those related to older populations (aged, elder, and older adult) and coaching (health or wellness coaching). This combination of search terms retrieved 225 papers relevant to health coaching. The review of these papers was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Figure 1) [23]. After deleting duplicates, 196 papers were included for an abstract review and screening. These abstracts were reviewed for studies that discussed health coaching for populations that included older adults (aged 65 and older) combined with some form of remote monitoring. The abstract screening yielded 42 articles for full-text review, of which 13 articles were identified that met the eligibility criteria (health coaching, remote monitoring, and older adults). Studies were excluded from our review if older populations (aged 65 and over) were not included, if the study did not include remote monitoring (RPM or RAM), or if the study did not include some form of coaching intervention. Subsequent to the review, 2 additional studies were identified [24,25] which provided the results for the ACTIVATE Trial [26] included in the original search. The results of the literature review were charted based on the following criteria: description of the coaching intervention, type of remote monitoring, study type, size of the study population, length of the study, condition monitored, and the outcomes.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart.



Results

The results of the literature review are summarized in Table 1. All 13 studies were published between 2014 and 2019. Four studies were randomized controlled trials that ranged from 83 to 1437 participants [12,26-28]. One study was a quasi-experiment (nonrandomized cohort study) with 144 participants [29]. Six studies were pilot trials that ranged from 6 to 33 participants [21,30-34]. There was 1 qualitative interview of 10 health care workers [35] and 1 user acceptance study with 11 participants [36]. The main goal of each of these studies was to evaluate the effectiveness of health coaching. Our review focused on the effectiveness of health coaching (human coach versus automated coaching system) combined with remote monitoring technology (RAM and RPM) for older adults.

Table 1. Summary of results from the literature review.

Study	Coaching intervention	Type of remote monitoring	Study type	Sample size, n	Study duration	Condition monitored	Outcomes
[12]	Human coach and telephone calls	RPM[b]	Randomized control trial	1437	26 weeks	Chronic heart failure	Re-admissions (N[e]), mortality (N), and quality of life (Y[f])
[27]	Human coach and telephone calls	RAM[c] and RPM	Randomized control trial	595	1 year	Chronic heart failure, diabetes	Blood pressure (N), body weight (N), and quality of life (N)
[26]	Human coach and telephone calls	RAM	Randomized control trial	83	12 weeks	Cancer	Physical activity (Y) and sedentary behavior (Y)
[28]	Human coach, telephone calls, and mobile app	RAM and RPM	Randomized control trial	131	26 weeks	Diabetes	HbA_{1c}[h] level (N), body weight (Y), and quality of life (N)
[29]	Human coach, telephone calls, SMS text messages, online training, and social networking	RAM and RPM	Quasi-experiment	144	1 year	Diabetes	HbA_{1c} (Y) and body weight (Y)
[30]	Human coach, DSS[a], telephone calls, and exercise videos	RAM and RPM	Pilot study	12	8 weeks	Chronic obstructive pulmonary disease	Program adherence (Y) and patient satisfaction (Y)
[31]	Human coach, telephone calls, and SMS text messages	RAM	Pilot study	24	4 weeks	Cancer	Physical activity (Y) and fatigue (Y)
[32]	Human coach, DSS, exercise videos, and SMS text messages	RAM	Pilot study	6	2-6 weeks	General health	Yes
[21]	Human coach, DSS, exercise videos, SMS text messages	RAM and RPM	Pilot study	33	Various	General health	Behavior change (I[g])
[33]	Human coach, telephone calls, and mobile app	RAM and RPM	Pilot study	21	26 weeks	Diabetes	HbA_{1c} level (Y) and body weight (Y)
[34]	Automated coach, DSS, SMS text messages, and mobile app	RAM	Pilot study	27	26 weeks	Diabetes	HbA_{1c} level (Y) and activity level (Y)
[35]	Human coach and SMS text messages	RAM	Qualitative interview	10	N/A[d]	General health	Inconclusive
[36]	Human coach, telephone calls, and mobile app	RAM and RPM	User acceptance study	11	26 weeks	Diabetes	User acceptance (Y)

[a]DSS: decision support system.

[b]RPM: remote patient monitoring.

[c]RAM: remote activity monitoring

[d]N/A: not applicable.

[e]N: not effective.

[f]Y: effective.

[g]I: inconclusive.

[h]HbA_{1c}: hemoglobin A_{1c}.

The predominate type of health coaching was via a human coach (12/13 studies) [12,21,26-33,35,36], whereas an automated health coaching system was employed in only 1 study [34]. Human coaching included an initial training session [12,24-26,28,31,36], periodic training sessions [29], scheduled periodic contact with patients [12,24-27,29-31,33,35,36], or

interventional contact based on remote monitoring results [12,21,28,32]. Four studies employed the use of a DSS to augment or assist the health coach [21,30,32,34]. The DSSs included software programs that generated trends and alerts for the health coach based on the remotely monitored data [30], artificial intelligence systems that evaluated the remotely monitored data and provided recommendations to the health coach [21,32], and a fully automated system that monitored physical activity and provided tailored feedback to the patient based on the monitored results [34]. Four studies employed the use of a mobile app for remote monitoring [28,33,34,36]. RAM was the most common type of telehealth technology employed (12 studies) [21,26-36] followed by RPM (8 studies) [12,21,27-30,33,36]. Communication with the patient was via telephone only (7 studies) [12,26-28,30,33,36], SMS text messages only (4 studies) [21,32,34,35], or telephone and SMS text messages (2 studies) [29,31]. Study durations ranged from 2 weeks to 1 year with 6 studies lasting 26 weeks or longer. The conditions monitored included diabetes (6 studies) [27-29,33,34,36] cancer (2 studies) [26,31], chronic heart failure (2 studies) [12,27], chronic obstructive pulmonary disease (1 study) [30], and overall general health (3 studies) [21,32,35].

Effectiveness outcomes assessed included hospital admissions/re-admissions, mortality, hemoglobin A_{1c} (HbA_{1c}) level, body weight, blood pressure, physical activity level, fatigue, quality of life, and user acceptance of the coaching program and technology. Of the 13 studies reviewed, 10 found coaching supported by telehealth technology to be effective in at least one of the outcomes assessed in the studies [12,26,28-34,36]. As much as 5 of the 6 studies that monitored diabetes found health coaching plus remote monitoring to be effective particularly for physical activity level and body weight [28,29,33,34,36]. Neither of the 2 studies that monitored chronic heart failure found health coaching plus remote monitoring to be effective [12,27] except for improving one's quality of life in one of the studies [12]. Both studies that monitored patients with cancer found health coaching plus remote monitoring to be effective at improving the patient's physical activity level [26,31]. Only 1 [32] of the 3 studies that monitored general health [21,32,35] found health coaching plus remote monitoring to be effective. In summary, the results indicate that health coaching plus remote monitoring can be effective at improving a patient's physical activity level, HbA_{1c} values, and in reducing body weight.

Discussion

Principal Findings

Health coaching that incorporates telehealth technologies has been implemented in older populations with mixed results. As much as 10 of the 13 studies reviewed found this method of health coaching to provide effective outcomes [12,26,28-34,36]. This literature review identified several gaps that warrant discussion or additional research.

Human Versus Automated Coach

One of the more prominent findings identified in this review was the dependence on a human to provide health coaching and interaction with the patient. As much as 12 of the 13 studies reviewed included a human coach [12,21,26-33,35,36], and thus the outcomes were probably heavily reliant on a human in the process. Four of the studies did include health coaching systems that incorporated the use of a DSS [21,30,32,34]; however, only 1 study completely replaced the human coach with a DSS [34].

The health coaching system in the Yom-Tov et al's pilot study [34] was fully automated in that neither the patient nor the health coach had to manually enter data or actions into the DSS or remote monitoring system after the patient's activity goals were established. A smartphone app recorded the patient's physical activity and transmitted the data to the DSS. A tailored daily feedback SMS text message was sent to each participant to encourage exercise. An algorithm determined the message to be sent based on whether the patient reached his/her activity goal the previous day. The study found that customizing or changing the daily message based on the actual physical activity performed was effective at getting the patient to increase daily activity whereas a constant daily reminder message was not effective. The use of a DSS to augment or replace human coaching indicates there is some movement toward augmenting the human coach with DSS technology. A benefit of using a DSS combined with remote monitoring is the ability to provide 24/7 continuous monitoring and intervention which may not be possible with a human coach. Although costs were not assessed in these studies, it is surmised that lessening the amount of direct human involvement in the coaching process should reduce overall cost. Additional studies should be performed with the focus of comparing the clinical and cost-effectiveness of the following 3 forms of health coaching: (1) human health coach only, (2) health coaching performed by a DSS only, and (3) a hybrid model of health coaching by a human coach augmented by a DSS.

Telephone Versus Electronic Media Communications

Another finding identified in this review was the heavy reliance on the use of a telephone to communicate with patients. Nine of the studies used a telephone for delivering coaching with mixed effectiveness results (2 of these studies augmented telephone communications with SMS text messages) [12,26-31,33,36]. The other 4 studies used DSS messages, SMS text messages, or video messages in lieu of telephone calls, also with mixed effectiveness results [21,32,34,35]. These results indicate that coaching effectiveness may not be dependent on the method of communication with the patient. Additional studies should be performed to evaluate the effectiveness and acceptance of using electronic media to communicate with the patient instead of live telephone calls.

Use of Smartphone Apps

Four studies included the use of a smartphone app as part of the integrated telehealth solution [28,33,34,36] with positive results for 3 of these studies [33,34,36]. Only one of these studies specifically evaluated the acceptance of smartphone app technology by the patients [36]. A recent qualitative study interviewed 12 community-dwelling older adults (aged 65-78) and found that older adults were, in general, satisfied with using technology to help monitor and manage their health on a daily basis (albeit amid some fears that technology would replace

human contact) [37]. Thus, there appears to be an opportunity to expand the use of technology, such as smartphone apps, as part of a telehealth system for older adults.

Coachability of Patients

Although not explicitly evaluated in the studies, it is probable that the results of these studies were dependent on the willingness of the patient to accept health coaching. Some of the studies evaluated the willingness of the patient to accept health coaching as part of the inclusion criteria while other studies only included patients who expressed an interest in the study. Thus, it can be assumed that most of the studies were biased toward those patients who are coachable. An opportunity exists to explore the effectiveness of health coaching using telehealth technology for patients who are not coachable.

Limitations

This literature review was focused on studies that included older adults (aged 65 and older) in the population assessed. Studies that excluded older adults were not included in our review, so the results should not be extrapolated to general populations. Most of the coaching interventions reviewed in this study included a human coach who provided feedback to participants via telephone calls. This type of coaching depends on the effort of the human coach to provide an adequate type of coaching to the participant which may or may not include all aspects of a robust coaching program (goal setting, education, encouragement, and feedback on health-related behaviors). In addition, the studies reviewed did not attempt to assess the capability of a human coach versus an automated health coaching system to effect behavior change. Additional research is needed to make this assessment. There was only 1 fully automated coaching intervention study found in our review, so no conclusion can be drawn regarding the effectiveness of automated health coaching interventions. Additional research is needed in the area of automated health coaching. The search criteria for this review focused first on health and wellness coaching that was then further filtered on remote monitoring as an element of the coaching. Several other studies of telehealth might have included coaching but not as a focus of the study.

Conclusions

Four inter-related issues face the health care industry: (1) the increasing numbers and percentage of older adults, (2) chronic disease management among this older population, (3) the desire of older adults to age in place, and (4) the cost of health care for older adults. Health coaching combined with telehealth technology has been shown to provide effective outcomes in 10 of 13 studies reviewed. Four studies included the use of a DSS to augment or replace the health coach with positive results. However, insufficient evidence of automated health coaching was found in our review to draw a conclusion regarding the efficacy of automated coaching. Although not assessed in these studies, the inclusion of automation in the health coaching process has the potential to reduce overall health care costs for older adults. The benefits of health coaching combined with telehealth are evident and should be further explored.

Future Directions

One of the more prominent findings identified in this review was the dependence on a human to provide health coaching and interaction with the patient. Thus, the outcomes were probably heavily reliant on a human in the process. Future studies need to assess the capability of automated coaching systems versus human coaches to affect health behavior changes. Another prominent finding was the use of live telephone calls to provide coaching to the patient. Future studies should be performed to evaluate the effectiveness and acceptance of using electronic media to communicate with the patient. The studies reviewed did not specifically evaluate coachability or the willingness of the patient to accept health coaching. An opportunity exists to explore the effectiveness of health coaching using telehealth technology for patients who are not coachable. This discrepancy should be investigated by including quality of life measures in future studies of coaching systems. As sensors for RPM and RAM become more advanced and affordable, much more data will be available to monitor and evaluate. With advances in big data analytics, DSSs will be better informed and able to identify interventions when necessary. Based on the results of this review, additional studies should be conducted of the expanded use of health coaching and DSSs as part of the health care solution for older adults. In addition, cost-effectiveness of health coaching combined with telehealth needs to be assessed against human-only health coaching methods. The results of these studies would inform the future direction of health coaching.

Acknowledgments

The lead author of this literature review (CM) is a qualified veteran who receives funding (tuition and fees) from the Texas Veterans Commission through the Hazlewood Act. There is no involvement by the Texas Veterans Commission in the review and approval of this study.

Conflicts of Interest

None declared.

References

1. About Chronic Disease. Atlanta, GA: Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/chronicdisease/about/index.htm [accessed 2018-09-21]

2. Kochanek KD, Murphy SL, Xu J, Arias E. Deaths: Final Data for 2017. Natl Vital Stat Rep 2019 Jun;68(9):1-77 [FREE Full text] [Medline: 32501199]
3. Centers for Disease Control and Prevention. The state of aging and health in America 2013. Atlanta, GA: Centers for Disease Control and Prevention/U.S. Department of Health and Human Services; 2013. URL: https://www.cdc.gov/aging/pdf/state-aging-health-in-america-2013.pdf [accessed 2020-01-11]
4. De Nardi M, French E, Jones JB, McCauley J. Medical spending of the U.S. elderly (Report No.: 21270). Cambridge, MA: National Bureau of Economic Research; 2015 Jun. URL: https://www.nber.org/system/files/working_papers/w21270/w21270.pdf [accessed 2020-01-11]
5. How do health expenditures vary across the population? Peterson-KFF Health Syst Tracker. URL: https://www.healthsystemtracker.org/chart-collection/health-expenditures-vary-across-population/ [accessed 2020-05-31]
6. Colby SL, Ortman JM. Projections of the size and composition of the US population: 2014 to 2060. Suitland-Silver Hill, MD: U.S. Census Bureau; 2017. URL: http://wedocs.unep.org/bitstream/handle/20.500.11822/20152/colby_population.pdf [accessed 2021-01-11]
7. Khalfani-Cox L. What are the costs of aging in place?. 2017 Feb. URL: https://www.aarp.org/money/budgeting-saving/info-2017/costs-of-aging-in-place.html [accessed 2021-01-11]
8. Satariano WA, Scharlach AE, Lindeman D. Aging, place, and technology: toward improving access and wellness in older populations. J Aging Health 2014 Dec;26(8):1373-1389 [FREE Full text] [doi: 10.1177/0898264314543470] [Medline: 25502245]
9. Nathan V, Paul S, Prioleau T, Niu L, Mortazavi BJ, Cambone SA, et al. A Survey on Smart Homes for Aging in Place: Toward Solutions to the Specific Needs of the Elderly. IEEE Signal Process. Mag 2018 Sep;35(5):111-119. [doi: 10.1109/msp.2018.2846286]
10. Wijesooriya NR, Mishra V, Brand PLP, Rubin BK. COVID-19 and telehealth, education, and research adaptations. Paediatr Respir Rev 2020 Sep;35:38-42 [FREE Full text] [doi: 10.1016/j.prrv.2020.06.009] [Medline: 32653468]
11. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
12. Ong MK, Romano PS, Edgington S, Aronow HU, Auerbach AD, Black JT, Better Effectiveness After Transition–Heart Failure (BEAT-HF) Research Group. Effectiveness of Remote Patient Monitoring After Discharge of Hospitalized Patients With Heart Failure: The Better Effectiveness After Transition -- Heart Failure (BEAT-HF) Randomized Clinical Trial. JAMA Intern Med 2016 Mar;176(3):310-318 [FREE Full text] [doi: 10.1001/jamainternmed.2015.7712] [Medline: 26857383]
13. Kivelä K, Elo S, Kyngäs H, Kääriäinen M. The effects of health coaching on adult patients with chronic diseases: a systematic review. Patient Educ Couns 2014 Nov;97(2):147-157. [doi: 10.1016/j.pec.2014.07.026] [Medline: 25127667]
14. Oliveira JS, Sherrington C, Amorim AB, Dario AB, Tiedemann A. What is the effect of health coaching on physical activity participation in people aged 60 years and over? A systematic review of randomised controlled trials. Br J Sports Med 2017 Oct;51(19):1425-1432 [FREE Full text] [doi: 10.1136/bjsports-2016-096943] [Medline: 28320732]
15. Center for Connected Health Policy. About telehealth. 2019. URL: https://www.cchpca.org/about/about-telehealth [accessed 2020-01-02]
16. Hall A, Stellefson M, Bernhardt J. Healthy Aging 2.0: the potential of new media and technology. Prev Chronic Dis 2012;9:E67 [FREE Full text] [Medline: 22405474]
17. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
18. Bratan T, Clarke M. Towards the design of a generic systems architecture for remote patient monitoring. Conf Proc IEEE Eng Med Biol Soc 2005;2006:106-109. [doi: 10.1109/IEMBS.2005.1616353] [Medline: 17282122]
19. Mahoney DF. An Evidence-Based Adoption of Technology Model for Remote Monitoring of Elders' Daily Activities. Ageing Int 2010 Sep 23;36(1):66-81 [FREE Full text] [doi: 10.1007/s12126-010-9073-0] [Medline: 21423843]
20. Welch G, Balder A, Zagarins S. Telehealth program for type 2 diabetes: usability, satisfaction, and clinical usefulness in an urban community health center. Telemed J E Health 2015 May;21(5):395-403. [doi: 10.1089/tmj.2014.0069] [Medline: 25748544]
21. Pavel M, Jimison HB, Korhonen I, Gordon CM, Saranummi N. Behavioral Informatics and Computational Modeling in Support of Proactive Health Management and Care. IEEE Trans Biomed Eng 2015 Dec;62(12):2763-2775 [FREE Full text] [doi: 10.1109/TBME.2015.2484286] [Medline: 26441408]
22. Jonk Y, Lawson K, O'Connor H, Riise KS, Eisenberg D, Dowd B, et al. How effective is health coaching in reducing health services expenditures? Med Care 2015 Feb;53(2):133-140 [FREE Full text] [doi: 10.1097/MLR.0000000000000287] [Medline: 25588134]
23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
24. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. Maintenance of physical activity and sedentary behavior change, and physical activity and sedentary behavior change after an abridged intervention: Secondary

outcomes from the ACTIVATE Trial. Cancer 2019 Aug 15;125(16):2856-2860 [FREE Full text] [doi: 10.1002/cncr.32142] [Medline: 31012968]
25. Lynch BM, Nguyen NH, Moore MM, Reeves MM, Rosenberg DE, Boyle T, et al. A randomized controlled trial of a wearable technology-based intervention for increasing moderate to vigorous physical activity and reducing sedentary behavior in breast cancer survivors: The ACTIVATE Trial. Cancer 2019 Aug 15;125(16):2846-2855. [doi: 10.1002/cncr.32143] [Medline: 31012970]
26. Lynch BM, Nguyen NH, Reeves MM, Moore MM, Rosenberg DE, Wheeler MJ, et al. Study design and methods for the ACTIVity And TEchnology (ACTIVATE) trial. Contemp Clin Trials 2018 Jan;64:112-117 [FREE Full text] [doi: 10.1016/j.cct.2017.10.015] [Medline: 29097298]
27. Karhula T, Vuorinen A, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and Mobile Phone-Based Health Coaching Among Finnish Diabetic and Heart Disease Patients: Randomized Controlled Trial. J Med Internet Res 2015;17(6):e153 [FREE Full text] [doi: 10.2196/jmir.4059] [Medline: 26084979]
28. Wayne N, Perez DF, Kaplan DM, Ritvo P. Health Coaching Reduces HbA1c in Type 2 Diabetic Patients From a Lower-Socioeconomic Status Community: A Randomized Controlled Trial. J Med Internet Res 2015;17(10):e224 [FREE Full text] [doi: 10.2196/jmir.4871] [Medline: 26441467]
29. Sepah SC, Jiang L, Peters AL. Translating the Diabetes Prevention Program into an Online Social Network: Validation against CDC Standards. Diabetes Educ 2014 Apr 10;40(4):435-443. [doi: 10.1177/0145721714531339] [Medline: 24723130]
30. Benzo RP, Kramer KM, Hoult JP, Anderson PM, Begue IM, Seifert SJ. Development and Feasibility of a Home Pulmonary Rehabilitation Program With Health Coaching. Respir Care 2018 Feb;63(2):131-140 [FREE Full text] [doi: 10.4187/respcare.05690] [Medline: 29066590]
31. Gell NM, Grover KW, Humble M, Sexton M, Dittus K. Efficacy, feasibility, and acceptability of a novel technology-based intervention to support physical activity in cancer survivors. Support Care Cancer 2017 Apr;25(4):1291-1300. [doi: 10.1007/s00520-016-3523-5] [Medline: 27957621]
32. Jimison H, Hagler S, Kurillo G, Bajcsy R, Pavel M. Remote health coaching for interactive exercise with older adults in a home environment. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5485-5488 [FREE Full text] [doi: 10.1109/EMBC.2015.7319633] [Medline: 26737533]
33. Wayne N, Ritvo P. Smartphone-enabled health coach intervention for people with diabetes from a modest socioeconomic strata community: single-arm longitudinal feasibility study. J Med Internet Res 2014;16(6):e149 [FREE Full text] [doi: 10.2196/jmir.3180] [Medline: 24907918]
34. Yom-Tov E, Feraru G, Kozdoba M, Mannor S, Tennenholtz M, Hochberg I. Encouraging Physical Activity in Patients With Diabetes: Intervention Using a Reinforcement Learning System. J Med Internet Res 2017 Oct 10;19(10):e338 [FREE Full text] [doi: 10.2196/jmir.7994] [Medline: 29017988]
35. Brandt CJ, Søgaard GI, Clemensen J, Søndergaard J, Nielsen JB. Determinants of Successful eHealth Coaching for Consumer Lifestyle Changes: Qualitative Interview Study Among Health Care Professionals. J Med Internet Res 2018 Jul 05;20(7):e237. [doi: 10.2196/jmir.9791] [Medline: 29980496]
36. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: A qualitative inquiry. J Telemed Telecare 2015 Jul 21. [doi: 10.1177/1357633X15595178] [Medline: 26199275]
37. 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]

Abbreviations

ADL: activity of daily living
DSS: decision support system
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAM: remote activity monitoring
RPM: remote patient monitoring

Edited by A Kushniruk; submitted 23.08.20; peer-reviewed by D Chrimes, N Wayne, A Martínez-García, R Goodison, E Stringer; comments to author 18.10.20; revised version received 17.12.20; accepted 23.12.20; published 29.01.21.

Please cite as:
Markert C, Sasangohar F, Mortazavi BJ, Fields S
The Use of Telehealth Technology to Support Health Coaching for Older Adults: Literature Review
JMIR Hum Factors 2021;8(1):e23796
URL: http://humanfactors.jmir.org/2021/1/e23796/
doi:10.2196/23796
PMID:33512322

Original Paper

Perceptual Gaps Between Clinicians and Technologists on Health Information Technology-Related Errors in Hospitals: Observational Study

Theophile Ndabu[1*], MBA; Pavankumar Mulgund[1*], PhD; Raj Sharman[1*], PhD; Ranjit Singh[2*], MBA, MD

[1]Department of Management Science and Systems, School of Management, State University of New York at Buffalo, Buffalo, NY, United States
[2]School of Medicine and Biomedical Sciences, State University of New York at Buffalo, Buffalo, NY, United States
[*]all authors contributed equally

Corresponding Author:
Theophile Ndabu, MBA
Department of Management Science and Systems
School of Management
State University of New York at Buffalo
203 Alfiero Center
Buffalo, NY, 14260
United States
Phone: 1 7166453271
Email: tnndabu@buffalo.edu

Abstract

Background: Health information technology (HIT) has been widely adopted in hospital settings, contributing to improved patient safety. However, many types of medical errors attributable to information technology (IT) have negatively impacted patient safety. The continued occurrence of many errors is a reminder that HIT software testing and validation is not adequate in ensuring errorless software functioning within the health care organization.

Objective: This pilot study aims to classify technology-related medical errors in a hospital setting using an expanded version of the sociotechnical framework to understand the significant differences in the perceptions of clinical and technology stakeholders regarding the potential causes of these errors. The paper also provides some recommendations to prevent future errors.

Methods: Medical errors were collected from previous studies identified in leading health databases. From the main list, we selected errors that occurred in hospital settings. Semistructured interviews with 5 medical and 6 IT professionals were conducted to map the events on different dimensions of the expanded sociotechnical framework.

Results: Of the 2319 identified publications, 36 were included in the review. Of the 67 errors collected, 12 occurred in hospital settings. The classification showed the "gulf" that exists between IT and medical professionals in their perspectives on the underlying causes of medical errors. IT experts consider technology as the source of most errors and suggest solutions that are mostly technical. However, clinicians assigned the source of errors within the people, process, and contextual dimensions. For example, for the error "Copied and pasted charting in the wrong window: Before, you could not easily get into someone else's chart accidentally...because you would have to pull the chart and open it," medical experts highlighted contextual issues, including the number of patients a health care provider sees in a short time frame, unfamiliarity with a new electronic medical record system, nurse transitions around the time of error, and confusion due to patients having the same name. They emphasized process controls, including failure modes, as a potential fix. Technology experts, in contrast, discussed the lack of notification, poor user interface, and lack of end-user training as critical factors for this error.

Conclusions: Knowledge of the dimensions of the sociotechnical framework and their interplay with other dimensions can guide the choice of ways to address medical errors. These findings lead us to conclude that designers need not only a high degree of HIT know-how but also a strong understanding of the medical processes and contextual factors. Although software development teams have historically included clinicians as business analysts or subject matter experts to bridge the gap, development teams will be better served by more immersive exposure to clinical environments, leading to better software design and implementation, and ultimately to enhanced patient safety.

(JMIR Hum Factors 2021;8(1):e21884) doi:10.2196/21884

KEYWORDS

patient safety; medical errors; health information technology; sociotechnical framework; patient harm

Introduction

Background

The widespread use of information technology (IT) has contributed to improved patient safety in the hospital setting [1-5]. However, many different kinds of medical errors attributable to the use of IT in health care have negatively impacted patient safety [6,7]. The number of patients who experience adverse events is estimated to be 40% of all patients who visit primary and ambulatory care [8]. These safety events may lead to an extended hospital stay, additional side effects, or distress and in some cases death. In addition to the loss of life and health impairment, the consequences of adverse events include increased financial costs to patients and the society at large [9].

In hospital settings, several benefits, including health care delivery improvement and reduction in medication errors, have been attained through the use of health information technology (HIT) [3]. However, new patient safety errors attributable to the use of HIT continue to be a significant issue [7]. For example, according to a recent study [10], in Pennsylvania alone, a total of 889 medication error reports listed HIT as a factor contributing to events submitted to the Pennsylvania Patient Safety Authority in the first 6 months of 2016. The study also shows that dose omission, wrong dosage, and extra dosage were the most commonly reported events. The most common HIT systems implicated in the events were the computerized prescriber order entry system, the pharmacy system, and the electronic medication administration record. Several government agencies and academic and clinical practitioner committees have been concerned about the unintended consequences of introducing IT in clinical environments. Several articles [9-11] report such adverse patient safety events related to HIT and emphasize the need for more cohesive HIT development processes to reduce the gulf of evaluation between medical and IT teams.

This pilot study seeks to classify patient safety events in hospital settings and to understand the differing perspectives of HIT designers and users concerning the potential causal factors of technology-related medical errors. In addition, the study suggests prescriptive measures to prevent reoccurrences of errors. Understanding the perspectives of both medical and IT stakeholders could help resolve the root causes of medical errors. The proposed classification could be used in facilitating medical and technology stakeholders in working together and working through different perspectives on the causes of HIT-related errors to identify likely solutions and ultimately design better HIT artifacts. To better understand the significant differences, we selected from our list of errors collected through the literature review, 12 archetype errors that occurred in a clinical setting, and examined them using the lens of sociotechnical theory from both clinical and IT systems perspectives. In the next section, we introduce the sociotechnical framework and present the proposed error classification. Following this, the Methods section details data collection and analysis. Subsequently, the results and discussion are presented before the Conclusions section.

Sociotechnical Framework

The sociotechnical theory posits that organizational performance depends on the interactions between social and technical factors, grouped into 4 pillars: technology, process, people, and environment [12]. Prior research suggests that developing applications that cater to end-user needs requires designers and developers to understand the workflow structures, organizational culture, and environment in which these systems will operate [13]. Hence, patient safety improvement is contingent on the joint optimization of social and technical factors in the hospital setting.

This paper creates a more detailed taxonomy by adding subcomponents of the 4 central pillars to the sociotechnical framework [12,13]. The expanded taxonomy allows for a better classification of errors and the development of more precise solutions. Furthermore, we classify the errors in terms of the causes based on the feedback of medical experts and IT professionals. Using the results of this classification process, we provide more in-depth insights into the significant differences in medical and clinical staff members' and IT professionals' perceptions regarding these errors and offer a prescription to mitigate them.

Several studies have used the sociotechnical framework to examine several aspects of HIT implementation and use, including human-computer interaction [14], the impact of policy, infrastructure, and people on the quality of health information [15], ergonomic and macroergonomic aspects of health technologies [16-20], risk assessment of electronic medical record safety [18], and usability factors [14,18]. The sociotechnical framework has also been used to classify patient safety events [21-23]. However, these studies have classified errors on the sociotechnical framework's high-level dimensions on which errors map the most (Table 1 shows a comparison of the 3 published papers closest to our efforts and details how this study is different). The sociotechnical framework suggests that multiple forces from multiple dimensions (and different hierarchical levels of a particular dimension) are at work when errors occur [24]. As patient safety events occur in a complex environment, there is a need for a classification that considers the impacts of multiple dimensions of the framework on each patient safety event's occurrence. Table 1 provides a summary differentiating the studies closest to the work in this paper. These studies were included because the authors used the sociotechnical framework to classify medical errors [21,23] or HIT-related sentinel events [22].

Table 1. A comparison with previous studies based on the use of the sociotechnical framework

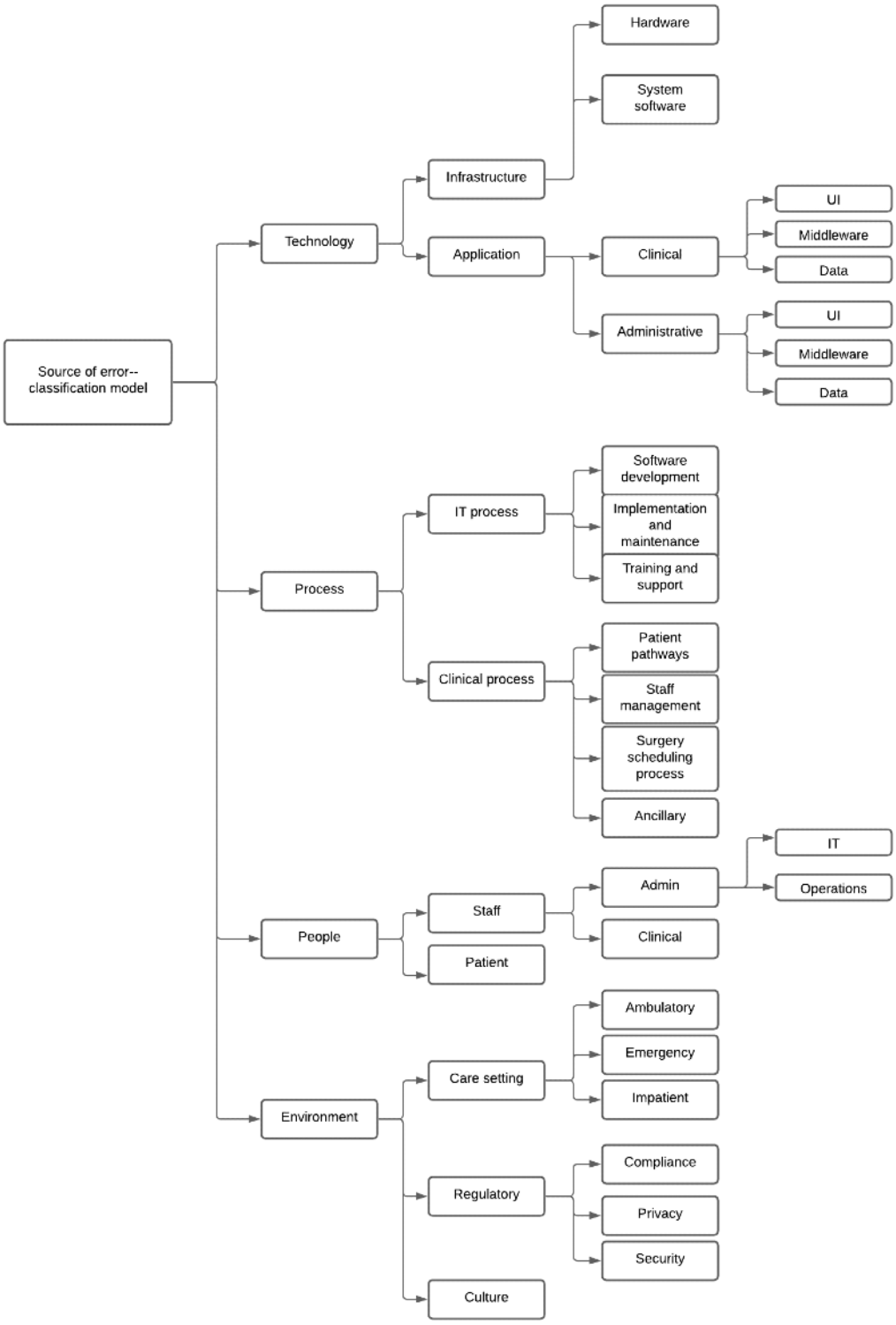
Studies (references)	Methodologies for error classification				
	Errors classified in 1 high-level dimension only—fitting one dimension excludes others	Errors classified in one dimension and its subdimensions only—fitting one dimension excludes others	Errors classified in multiple high-level dimensions	Classification based on multiple dimensions and their subcomponents	One error at a time
Safety huddles to proactively identify and address electronic health record safety [21]	✓[a]	✓	—[b]	✓	✓
Contribution of sociotechnical factors to health information technology–related sentinel events [22]	✓	—	—	—	—
Exploring the sociotechnical intersection of patient safety and electronic health record implementation [23]	✓	—	—	—	—
This study	—	—	✓	✓	✓

[a]Methodology applicable to the study.

[b]Methodology not applicable to the study.

Medical error classifications have been developed using other approaches. The System Theoretic Accidents Models and Process framework has been used to classify medical errors in 3 broad categories: feedback, control action, and knowledge errors [25]. The Human Factors Classification Framework [26] has been adapted to health care to classify medical errors in 5 categories: decision errors, skill-based errors, perceptual errors, routine violations, and exceptional violations [27,28]. Other studies have developed taxonomies without the use of a particular framework [29-31]. Prior studies have not applied the sociotechnical framework on medical errors with the intent of exploring the root causes and potential avenues through which the errors can be fixed. Furthermore, the dimensions of sociotechnical frameworks described in the extant research literature have not considered the emergence of new technologies such as cloud computing, n-tier architectures, and new management paradigms, including DevOps and microservices architecture. We adapted and extended the sociotechnical framework with additional dimensions that reflect new trends in IT. A group of expert researchers in information systems and sociotechnical theory reviewed this model [32]. Feedback from these experts was incorporated to refine the classification model, which is presented in Figure 1.

Figure 1. Error classification model. UI: user interface.



Proposed Classification

Sociotechnical theory emphasizes the interplay of the social and technical aspects of adopting and using technology [17,18,33]. The theory hinges on four basic constructs (technology, people, process, and environment) and the interaction between these constructs. In the expanded version of the sociotechnical framework, we detail the components of the technology dimension to include the IT infrastructure, which in turn comprises hardware, software, and apps. These also include emerging technologies, such as cloud computing, the internet of things, mobile apps, and the use of artificial intelligence, predictive and prescriptive analytics, and robotics. The technology dimension can also be partitioned based on the type of use, broadly classified as either administrative (including administrative IT and resource scheduling) or clinical. The need to investigate at this level of detail stems from the fact that the type of interaction varies based on the interacting

subcomponents. Furthermore, the app layers can be viewed as comprising the user interface, middleware (including the logic layer), backend (including the logic layer), and data.

The process dimension includes administrative and clinical workflows. Administrative workflows related to IT include the collection, storage, processing, and presentation of information for more effective resource management, such as clinical and IT staff management, operating room scheduling, risk and safety management, billing and facility management, and inventory management to ensure the business management of the hospitals. The subdimensions of IT processes are software development, HIT implementation and maintenance, and training and support. Clinical processes include patient record management, clinical pathways, patient bed assignment, and physician notes. Some processes are both clinical and administrative; these include the inventory management of drugs and clinical supplies, surgery room and equipment scheduling, and patient discharge management. Processes in health care settings allow all stakeholders to perform tasks in a predetermined manner to obtain successful outcomes [24,34,35]. Patient safety errors manifest when there is a misalignment between the elements of IT and clinical processes.

The people dimension includes patients, clinical staff, and administrative staff. People interact with each other and with the technology available to them. The hospital employee space consists of providers with different competencies and clinical authorities and administrative staff with priorities that are often very different from those of clinical providers. Several examples are worth mentioning here. First, clinical staff members prioritize patients' clinical health, whereas IT personnel are more concerned with the processes involved in health care. Inconsistencies in their priorities often lead to errors. As people interact with the entire work system, a mismatch between people and any other components increases the risk of harm to patients. Human errors are also a threat to patient safety [36]. Therefore, it is essential to build user interfaces and systems that consider the priorities and goals of the different types of users of the system, and these goals go beyond the purely functional and technical requirements of the job.

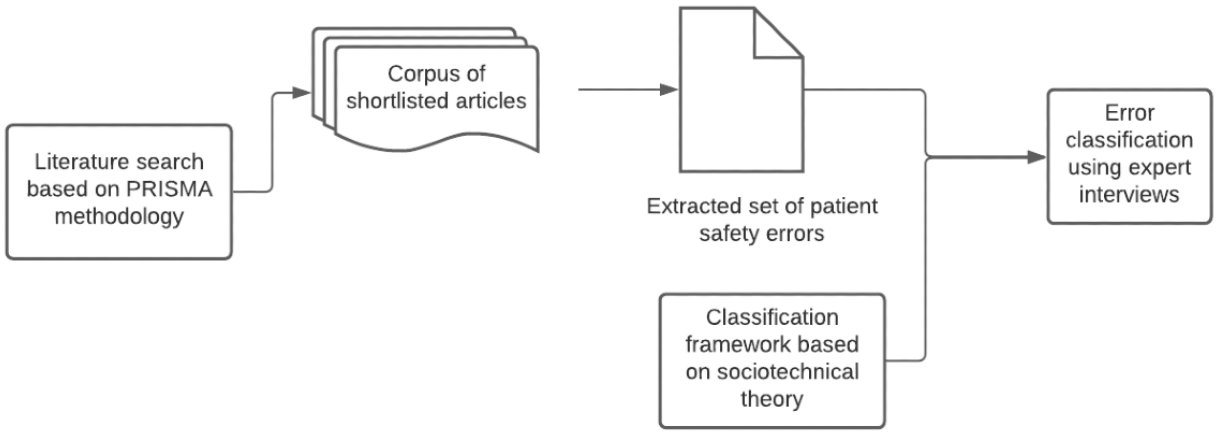
The environment consists of the care setting (eg, ambulatory, emergency, and in-patient), regulatory (eg, compliance, privacy, and security related), and culture. Culture stems from management style, organizational policy, and other systemic factors. Furthermore, different types of employees prioritize different goals, and conflicts in achieving these goals are often manifest in the building, implementation, and functioning of systems. Patients receiving services are external to the health care organization. To ensure more effective health care service provisioning, patient participation in the process is very important. In some areas, tasks must be performed by patients away from the health care organization. Contextual environments and skills to perform the required tasks differ from those of health care providers [33,35]. Regulations can also have a constraining effect on the error-free functioning of all subsystems. A thorough classification of patient safety events should consider specific areas of interaction between the environment dimension and all other dimensions. We use this expanded classification model to understand the gap in the mental models of clinical staff and technology professionals regarding the root cause of errors and how they should be addressed. We articulate our research design in the next section.

Methods

Research Design

The research design is comprised of 2 significant steps: developing a shortlisted set of IT-related patient safety issues and the classification of the root causes of medical errors with the sociotechnical lens using expert interviews. Figure 2 depicts the flow of the study.

Figure 2. Research flow.



Error Collection Using Literature Review

In this study, we first developed an extended sociotechnical framework that includes a finer level of granularity. Next, we systematically reviewed the literature on patient safety and medical errors from Ovid-MEDLINE, Embase, and Web of Science, which are leading medical databases in addition to Google Scholar. The systematic review process shown in Figure 2 aligns with commonly used steps of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses)

guidelines [37], as depicted by several exemplar papers [38-40]. The searches were performed using the following search terms: ("Patient Safety" OR "Medical") AND ("issue" OR "error") AND ("health information technology" OR "information technology"). Initially, the title, abstract, and index terms were used to screen published journal papers, conference papers, proceedings, case studies, and book chapters. We also used ancestral search to locate potentially relevant articles. Subsequently, the shortlisted papers were reviewed entirely. Two reviewers performed the screening independently. The reviewers met regularly to discuss the inclusion of the studies. A third reviewer was consulted when there was a discrepancy. Interrater reliability indicated a high agreement (Cohen κ value of 0.95).

Inclusion criteria included studies that addressed patient safety by identifying specific issues that occurred in health care settings and linked these errors to HIT. Furthermore, we excluded studies that were not available as the full text in the final search; were not in English; or were reports, abstracts only, letters, or commentaries.

Expert Interviews

An invitation email to participate in the study was sent to the alumni of the University at Buffalo. The email contained the eligibility criteria consisting of ≥5 years of HIT experience and at least 1 IT-related professional certification. A separate invitation email mentioning the selection criteria was sent to medical experts through the Office of Business Coordination at the University at Buffalo. A minimum experience of 5 years working as a medical doctor or as a registered nurse was required to qualify for the interview. All participants who responded met the selection criteria and were included in the study.

To better understand the perspectives of different stakeholders, we conducted multiple semistructured interviews [41] with different stakeholders, namely 6 IT and 5 medical experts to map the errors on the different dimensions of the expanded sociotechnical framework. Experts could map an error on multiple (or on all) subdomains of the sociotechnical framework to show the different sociotechnical factors that could contribute to the error. The purpose of accounting for the different perspectives was to understand how each group understood the predicates of the problem and allow us to reflect on how best the error could be addressed. Interviews were selected based on their domain experience, education, and industry certifications. The IT experts, who were recruited from the alumni list of the State University of New York at Buffalo, were software development professionals with a master's degree and IT professional certifications, such as the certified scrum master, the health level 7 control specialists, and the project management professional certifications. The minimum work experience cutoff for IT experts was 5 years for HIT in addition to possessing at least one IT-related professional certification.

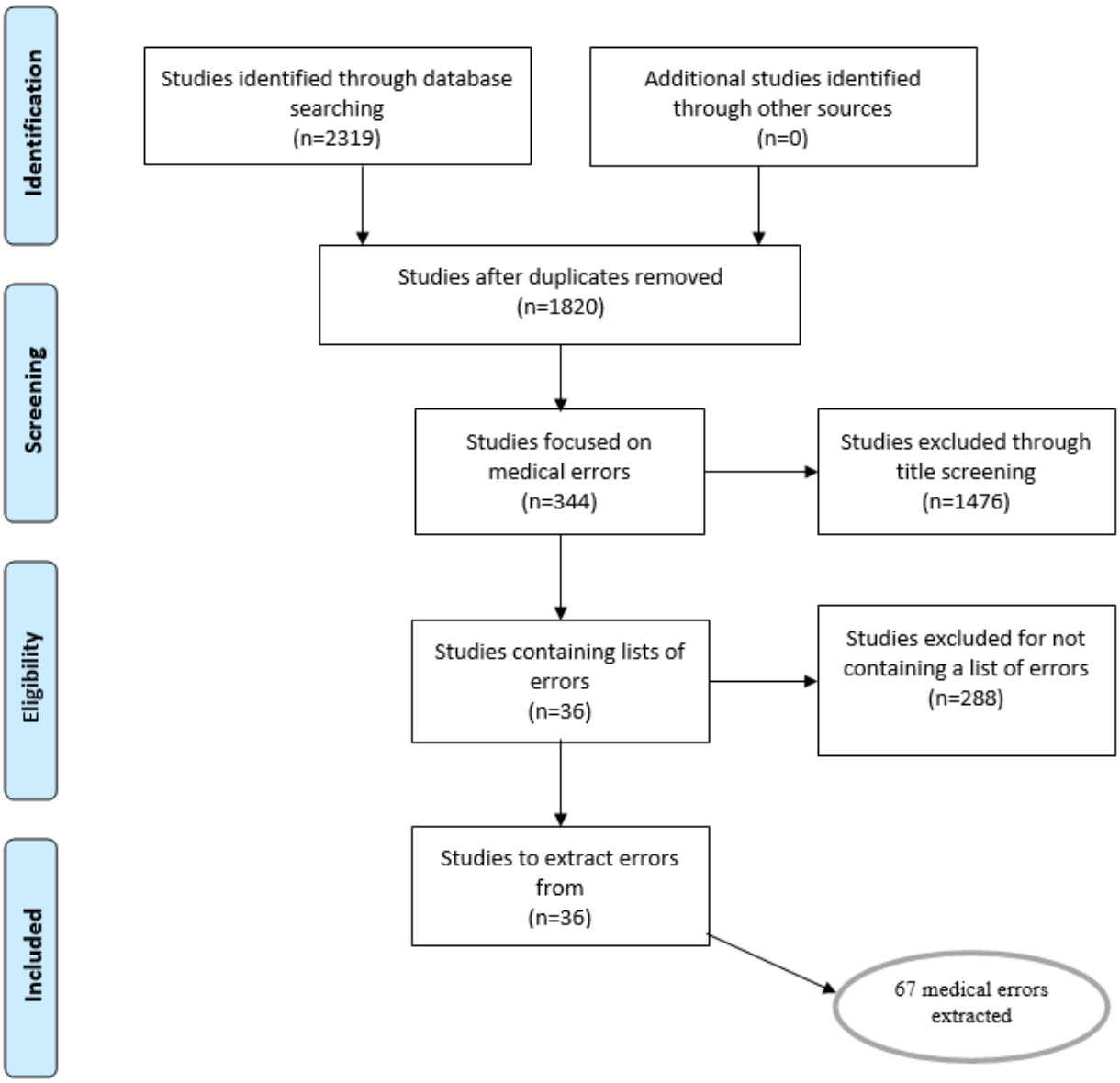
IT experts who were interviewed had extensive IT experience (mean 10.33, SD 1.11 years) with significant HIT experience (mean 8.83, SD 2.03 years; Multimedia Appendix 1 uploaded as a supplementary file for brief profiles of IT interviewees). The medical experts interviewed were physicians and registered nurses with broad primary care experience from working with multiple health care institutions across the United States and Canada. They are all currently working with hospitals and institutions affiliated with the university at Buffalo (Multimedia Appendix 2). Medical experts had a mean experience of 16.6 (SD 7.33) years. The minimum and maximum numbers of years of HIT experience for IT experts were 5 and 12, respectively. The work experience of medical experts varied from 8 to 27 years. The questionnaire and interview process are detailed in Multimedia Appendix 3. Experts were asked to provide their opinions on why the selected errors (Multimedia Appendix 4 [42-48]) occurred and how the errors could be prevented. The extensive experience of both IT and medical experts in their respective domains qualifies them to map medical errors on the sociotechnical framework. The study was approved in November 2019 (IRB# STUDY00003838).

Results

Search Results

The literature search resulted in 344 articles, 141 of which were duplicates. After removing articles based on their content, we retained 36 articles [10,28,42-47,49-76] that met the 2 criteria set for the study. We then extracted 67 unique patient safety events from the articles in which 12 specific issues related to IT use in the hospital setting were shortlisted. The process followed the PRISMA methodology [37] as detailed in Figure 3. The remaining errors occurred outside a health care setting and were excluded from the study. The error description includes the error context in the literature review format commonly known as problems, interventions, comparisons, and outcomes model [37]. The articles describing the errors contained a clear purpose, literature review, research methodology, results, and conclusions.

Figure 3. Data collection method.



Study Characteristics and Error Classification

In this study, experts categorized errors based on their opinion of where the source of the error lies. Experts were provided with the definitions of the elements of the framework and were informed that an error could result from multiple sources. They were asked to map each error at the lowest level of one or multiple dimensions of the sociotechnical framework. The authors then interacted with the experts to understand the reasons behind their mapping selection. The interactions included questions related to suggestions on the best way to address the problems and prevent them from occurring. In line with extant literature on data analysis in qualitative research coding [77,78], expert interviews were subsequently deconstructed into keywords and phrases and then grouped into ideas and concepts. The output of the analysis is summarized in the "key observations" below, for example, in Error 1: "Copied and pasted charting in the wrong window: Before, you could not easily get into someone else's chart accidentally...because you would have to pull the chart and open it."

Medical experts highlighted several contextual issues, such as the number of patients a health care provider is set to see in a short time frame, unfamiliarity with a new electronic medical record system, nurse transitions around the time of the error, and confusion due to patients having the same name. They emphasized process controls, including failure modes, as a potential fix. The technology experts discussed the lack of notification, poor user interface, and lack of end-user training as critical factors in this error. Error 2: "Incompatible data standards across multiple mobile applications led to the missing of vital data fields, which led to information loss."

Like the first sample, medical experts attributed this error to system software–related interoperability issues. They also highlighted several changes in the International Classification of Diseases (ICD) during the transition from ICD 9 to ICD 10 as an example of a situation that could lead to errors.

Technology experts, however, emphasized data formats, data transfer protocols, and service-orientated architecture as potential causes of errors.

Although we have detailed 2 instances here, the experts reviewed all 12 errors and identified the most likely set of possible dimensions to which the errors could be attributed. The sample errors used in the study are presented in Multimedia Appendix 2, and the results of analyzing these data are presented in Table 2, followed by several key observations.

Table 2. Classification by medical and IT experts.

Errors	Classification by medical experts	Classification by IT[a] experts
Nurse was supposed to enter a prescription for Amoxicillin 250 mg PO q8h×7 days (21 dispensed). However, the nurse failed to change the default dosage amount and dispensed too much medication (30 dispensed)	• UI[b]-clinical app implementation and maintenance • Clinical staff	• UI-clinical app software development • Clinical staff
Copied and pasted charting in the wrong window: "Before, you could not easily get into someone else's chart accidentally...because you would have to pull the chart and open it"	• Clinical staff • Clinical app • Training	• Clinical staff • In-patient
In general practice ward, the doctor consulted a patient with another patient's records and prescribed medications according to the wrong records. The patient died the same day of taking it. No further details were available	• Clinical staff • Clinical UI • Clinical middleware	• Clinical UI • Implementation and maintenance • Staff-admin (IT)
The receptionist intended to alert the general practitioner via the practice software about a patient with chest pain but instead sent the message to himself. The patient later died from a myocardial infarction	• UI-patient pathways • Clinical staff	• UI-clinical app software development • Training and support • Patient pathways
A patient received only half of their usual quantity of blood pressure medication because a repeat prescription for the medication did not transfer to a new software system when the patient's historical records were migrated. Because they did not have enough medication the patient tried to stretch out the old dose by taking the medication on alternate days. The patient had a stroke but made a full recovery.	• Software-systems • Patient pathways • Patient	• Data-clinical • Software development • Implementation and maintenance
A child had a full body x-ray. Some of the images went missing from the archival system where they were digitized. The x-ray was repeated to acquire the missing images, re-exposing the child to high levels of radiation	• Software-systems • Patient pathways	• Data-clinical
A compound in high demand such as Rifampicin was not listed in the computerized physician order entry system. The consequence was that the physician could not order rifampicin.	• Data-clinical • Ancillary • In-patient • Culture	• Data-clinical • Software development • Staff-admin (operations) • Culture
When an update is made to the frequency field on an existing prescription, the frequency schedule ID is not simultaneously updated on new orders sent to the pharmacy via (application)	• Software-development • Clinical-people • Software-systems	• Data-clinical • Staff-admin (IT) • Software-systems
Monitoring and Eavesdropping on Patient Vital Signs by hacking into the packet transfer from an internet of things device to the central system	• Middleware • Maintenance • People-staff (operations) • Compliance • Security	• System software • Data-clinical • Software development • Compliance • Security
Vulnerabilities of the hospital's IOT devices were exploited to initiate a denial-of-service attack to bring down hospital's servers which disrupted normal functioning	• Hardware • Software • IT implementation • Compliance • Security	• Compliance • Security
Use of portable devices that are not password protected makes the patient record vulnerable to the invasion of privacy	• Data-clinical • Software-development • Maintenance • Compliance • Privacy	• System software • Software-development • Security
Incompatible data standards across multiple mobile applications led to the missing of vital data fields which led to information loss	• Software-systems • Software-development	• Data-clinical

[a]IT: information technology.

[b]UI: user interface.

Discussion

Principal Findings

Some of the crucial observations include (1) The identified potential sources of the errors and solution areas differed considerably between clinicians and IT specialists; (2) both groups identified multiple factors as potential causes of the errors; (3) the clinicians often focused on postproduction (eg, implementation, maintenance, training, context, and the way the application is used) issues as causal factors; (4) IT experts focused on software functionality, software development, and technical implementation issues as causal factors; (5) on most occasions when IT experts identified an issue as a "data" problem, clinicians seemed to think that the problem lay elsewhere, including the software system, software development, or patient pathways; (6) both groups seem to be congruent with the issues of compliance and security; and (7) IT experts rarely identified clinical pathways or workflows as an issue.

The classification of the identified medical errors using the framework is presented in Table 2. The continued occurrence of many errors is a reminder that current HIT software testing and validation do not seem adequate in terms of ensuring the functioning of the software within the health care organization. The attribution of the errors to different aspects of the sociotechnical framework by clinicians and IT professionals informs us that technologists and clinicians generally differ in their perspectives on factors that impact IT-related safety events. Software experts are often not acclimatized to the environment in which HIT software and tools are used, which could be a cause to the problem.

Although IT and medical experts' perceptions are similar in security and privacy, IT specialists often tend to assume that the issues are either software or hardware or user interface related. In contrast, clinicians tend to consider environmental, contextual, and process factors as contributors to patient safety events. The benefit of such a classification suggests that designers and developers who fix the errors consider the artifact's environment and the people using the artifact. A key realization is that such errors will continue to occur if health IT system developers do not fully grasp the importance of technology functioning in an environment of care delivery where the patient needs are paramount.

A careful review of the IT experts' classification of errors highlights the view that IT experts consider technology as the source of most errors and suggest solutions that are mostly technical. The IT experts highlighted the software systems and development as the top 2 sources of most errors. Similarly, the suggestions of potential fixes mostly revolve around the software. However, a common refrain that accompanied their answers was, "The doctor should double-check..." In contrast, clinicians tended to assign the source of errors within the people, process, and contextual (environmental) dimensions for the most part.

The difference in perspective could be explained by the fact that clinicians tend to deal with the system after implementation. In contrast, IT experts tend to look at the same problem from an IT development perspective. For example, for "Error 1," for which IT experts were asked how they would prevent a doctor from using the wrong chart when he had multiple charts open, the answer was always to restrict access to 1 open chart at a time. However, clinicians prefer having multiple windows open so that they can quickly consult with multiple patients in different rooms without having to close out and reopen a chart. For them, the issue is, "How easy is it for a physician to realize the mistake," and "Physicians should still be able to open multiple charts." The differing perspectives between designers and developers of the technology and its users can contribute to medical errors.

The development teams of clinical applications typically include clinicians who provide domain expertise. However, our study indicates that this may not be sufficient as IT experts do not fully grasp the clinical environment and how workloads and other patient-related variabilities impact the use of the software. Therefore, as a future investigation, we suggest that software companies immerse developers in clinical environments for a short period, so that the understanding of the environment is built into their psyche and translates into a more robust design.

HIT systems can be made less error prone if programmers and systems developers understand the health care organization's operating environment. Current systems do not have fail-safe mechanisms that could prevent some of the errors. For example, consider the documented error, "the nurse was supposed to enter a prescription...the nurse failed to change the default amount and dispensed too much medication"; from a software perspective, better checks and warnings can be developed. In this specific instance, a system challenge asking the nurse to review the dosing amount could have prevented the problem. From a process perspective, nurses could be trained to reexamine the dosage. Creating a poka-yoke (like a check-off box for dose amount) would force nurses to check the dosing before refilling the prescriptions. As the clinical experts and IT experts suggested slightly different predicates for the error, a solution that addresses the issue from both technical and from a process and workforce training perspective would provide multiple layers of defense against such failures. The different views expressed by IT and clinical experts can be used to create technical and process solutions so that there is a more robust defense against these types of errors.

Limitations and Future Studies

The results of this study should be interpreted cautiously, as there are several limitations to this study. The first shortcoming is related to the smaller number of participants interviewed in this study. Only 11 interviews comprising 5 medical providers and 6 HIT professionals were conducted. Therefore, this study should be considered a pilot study suggesting the differences in the mental models of the clinical and technical staff, which potentially leads to ineffective systems analysis and ultimately manifests as errors in practice. In addition, both IT and medical experts have, for the most part, acquired their education and expertise at affiliated institutions in the Northeast of the United States. Future studies should examine the hypothesis that medical experts are more likely to attribute medical errors to

contextual factors, whereas IT experts on technical factors use a nationally representative sample.

Second, we shortlisted 12 unique errors that occurred in a hospital setting; the findings of this study cannot be generalized beyond that context. Furthermore, we extracted the errors used in this study from articles written in the English language. Future studies could examine errors that occurred in medical homes, patients' homes, or other nonhospital settings or include studies written in other languages.

Third, the study did not examine errors that were discovered by HIT users before the occurrence of a patient safety event. Future studies should examine near-miss errors to determine their potential root causes and fixes using the lens of sociotechnical theory.

Conclusions

This study classifies medical errors gathered from extant literature based on an expanded sociotechnical framework. Interviews from health care and IT experts reveal differing perspectives on why medical errors occur in clinical settings. Health care experts were more likely to attribute the source of an error to the implementation and use of an IT tool, whereas IT experts were likely to identify software design and functionality as causal factors of medical errors. From the results of this study, we offer several error-prevention prescriptions that can be tested in future research. First, IT experts should observe the functioning of HIT postimplementation and collect metrics related to its impact on (1) physician consultation time, (2) physician efficiency, (3) patient-physician relationship, (4) training needs, and (5) how the software fits into the workflow and culture of the organization. Software developers should be trained to be sensitive to the provider and patient needs because their lack of exposure to postproduction issues and usage contexts leads to the development of applications that do not cater to all user situations. Understanding these situations may lead to building software constraints and improved user training. Although software development teams have historically included clinicians as business analysts or subject matter experts to bridge the gap, development teams will be better served by more immersive training and exposure to clinical environments, leading to better software design and software implementation strategies.

Authors' Contributions

TN, PM, R Sharman, R Singh contributed equally.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interviewees—information technology experts.
[DOCX File , 15 KB - humanfactors_v8i1e21884_app1.docx]

Multimedia Appendix 2
Interviewees—medical experts.
[DOCX File , 13 KB - humanfactors_v8i1e21884_app2.docx]

Multimedia Appendix 3
Interview process and questionnaire.
[DOCX File , 13 KB - humanfactors_v8i1e21884_app3.docx]

Multimedia Appendix 4
List of errors.
[XLSX File (Microsoft Excel File), 11 KB - humanfactors_v8i1e21884_app4.xlsx]

References

1. Buntin M, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471 [FREE Full text] [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
2. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 05;20(2):41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
3. Mishuris RG, Linder JA, Bates DW, Bitton A. Using electronic health record clinical decision support is associated with improved quality of care. Am J Manag Care 2014 Oct 01;20(10):445-452 [FREE Full text] [Medline: 25414982]
4. Jones S, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. Ann Intern Med 2014 Jan 07;160(1):48-54 [FREE Full text] [doi: 10.7326/m13-1531]

5. Menachemi N, Rahurkar S, Harle CA, Vest JR. The benefits of health information exchange: an updated systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1259-1265 [FREE Full text] [doi: 10.1093/jamia/ocy035] [Medline: 29718258]
6. Kim M, Coiera E, Magrabi F. Problems with health information technology and their effects on care delivery and patient outcomes: a systematic review. J Am Med Inform Assoc 2017 Mar 01;24(2):246-250 [FREE Full text] [doi: 10.1093/jamia/ocw154] [Medline: 28011595]
7. Sittig D, Wright A, Coiera E, Magrabi F, Ratwani R, Bates DW, et al. Current challenges in health information technology-related patient safety. Health Informatics J 2020 Mar;26(1):181-189 [FREE Full text] [doi: 10.1177/1460458218814893] [Medline: 30537881]
8. Auraaen A, Slawomirski L, Klazinga N. The economics of patient safety in primary and ambulatory care. OECD Health Working Papers 2018 Nov 29;106:1-57 [FREE Full text] [doi: 10.1787/baf425ad]
9. Van Den Bos J, Rustagi K, Gray T, Halford M, Ziemkiewicz E, Shreve J. The $17.1 billion problem: the annual cost of measurable medical errors. Health Aff (Millwood) 2011 Apr;30(4):596-603 [FREE Full text] [doi: 10.1377/hlthaff.2011.0084] [Medline: 21471478]
10. Lawes S, Grissinger M. Medication errors attributed to health information technology. Pennsylvania Patient Safety Advisory 2017 Mar;9(1):1 [FREE Full text]
11. Kohn L, Corrigan J. To err is human: building a safer health system. Washington, DC: National Academy Press 2000. [doi: 10.17226/9728] [Medline: 25077248]
12. Cooper R, Foster M. Sociotechnical systems. American Psychologist 1971;26(5):467-474. [doi: 10.1037/h0031539]
13. Berg M, Toussaint P. The mantra of modeling and the forgotten powers of paper: a sociotechnical view on the development of process-oriented ICT in health care. International J of Med Informatics 2003 Mar;69(2-3):223-234. [doi: 10.1016/s1386-5056(02)00178-8]
14. Sittig D, Singh H. A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 2010 Oct;19 Suppl 3:68-74 [FREE Full text] [doi: 10.1136/qshc.2010.042085] [Medline: 20959322]
15. Craig S, Kodate N. Understanding the state of health information in Ireland: A qualitative study using a socio-technical approach. Int J Med Inform 2018 Jun;114:1-5. [doi: 10.1016/j.ijmedinf.2018.03.001] [Medline: 29673598]
16. Woods D, Hollnagel E. Joint cognitive systems: patterns in cognitive systems engineering. In: Cognitive Ergonomics. Boca Raton, Florida: CRC Press; Mar 27, 2006.
17. Carayon P, Karsh B, Gurses AP, Holden RJ, Hoonakker P, Schoofs Hundt A, et al. Macroergonomics in health care quality and patient safety. Reviews of Human Factors and Ergonomics 2013 Sep 26;8(1):4-54. [doi: 10.1177/1557234x13492976]
18. Sittig D, Singh H. Toward more proactive approaches to safety in the electronic health record era. Jt Comm J Qual Patient Saf 2017 Oct;43(10):540-547 [FREE Full text] [doi: 10.1016/j.jcjq.2017.06.005] [Medline: 28942779]
19. Singh H, Upadhyay DK, Torretti D. Developing health care organizations that pursue learning and exploration of diagnostic excellence: an action plan. Acadamic Medicine 2020;95(8):1172-1178. [doi: 10.1097/acm.0000000000003062]
20. Rogers M, Sockolow PS, Bowles KH, Hand KE, George JE. Use of a human factors approach to uncover informatics needs of nurses in documentation of care. Int J Med Inform 2013 Nov;82(11):1068-1074 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.08.007] [Medline: 24008175]
21. Menon S, Singh H, Giardina TD, Rayburn WL, Davis BP, Russo EM, et al. Safety huddles to proactively identify and address electronic health record safety. J Am Med Inform Assoc 2017 Mar 01;24(2):261-267 [FREE Full text] [doi: 10.1093/jamia/ocw153] [Medline: 28031286]
22. Castro GM, Buczkowski L, Hafner JM. The contribution of sociotechnical factors to health information technology–related sentinel events. The Joint Commission Journal on Quality and Patient Safety 2016 Feb;42(2):70-AP3. [doi: 10.1016/s1553-7250(16)42008-8]
23. Meeks DW, Takian A, Sittig DF, Singh H, Barber N. Exploring the sociotechnical intersection of patient safety and electronic health record implementation. J Am Med Inform Assoc 2014 Feb 01;21(e1):28-34 [FREE Full text] [doi: 10.1136/amiajnl-2013-001762] [Medline: 24052536]
24. Donabedian A. The quality of medical care. Science 1978 May 26;200(4344):856-864. [doi: 10.1126/science.417400] [Medline: 417400]
25. Mason-Blakley F, Habibi R, Weber J, Price M. Assessing stamp EMR with electronic medical record related incident reports: case study: manufacturer and user facility device experience database. In: Proceedings of 2017 IEEE International Conference on Healthcare Informatics (ICHI). 2017 Presented at: 2017 IEEE International Conference on Healthcare Informatics; 23-26 Aug. 2017; Park City, UT, USA URL: https://doi.org/10.1109/ICHI.2017.97 [doi: 10.1109/ichi.2017.97]
26. Feyer A, Williamson AM. The role of work practices in occupational accidents. In: Proceedings of the Human Factors Society Annual Meeting. 2016 Aug 08 Presented at: Human Factors and Ergonomics Society Annual Meeting Proceedings p. 1100-1104 URL: https://doi.org/10.1177/154193129103501517 [doi: 10.1177/154193129103501517]
27. Mitchell R, Williamson A, Molesworth B. Application of a human factors classification framework for patient safety to identify precursor and contributing factors to adverse clinical incidents in hospital. Appl Ergon 2016 Jan;52:185-195 [FREE Full text] [doi: 10.1016/j.apergo.2015.07.018] [Medline: 26360210]

28. Mitchell R, Williamson A, Molesworth B. Use of a human factors classification framework to identify causal factors for medication and medical device-related adverse clinical incidents. Safety Science 2015 Nov;79:163-174 [FREE Full text] [doi: 10.1016/j.ssci.2015.06.002]
29. Taib I, McIntosh AS, Caponecchia C, Baysari MT. A review of medical error taxonomies: a human factors perspective. Safety Science 2011 Jun;49(5):607-615 [FREE Full text] [doi: 10.1016/j.ssci.2010.12.014]
30. Zhang J, Patel VL, Johnson TR, Shortliffe EH. A cognitive taxonomy of medical errors. J Biomed Inform 2004 Jun;37(3):193-204 [FREE Full text] [doi: 10.1016/j.jbi.2004.04.004] [Medline: 15196483]
31. Doumbouya MB, Kamsu-Foguem B, Kenfack H, Foguem C. Argumentative reasoning and taxonomic analysis for the identification of medical errors. Engineering Applications of Artificial Intelligence 2015 Nov;46:166-179. [doi: 10.1016/j.engappai.2015.08.009]
32. Ndabu T, Sharman R, Mulgund P, Garg R. Design principles for healthcare applications: a patient safety focus. 2019 Jun 04 Presented at: Design Science Research in Information Systems and Technology; 2019, June 4; Worcester, Massachussetts URL: https://drive.google.com/file/d/1Iu2SpWllKF3OsjZrBTSs9RACTJhNzz6y/view
33. Kushniruk A. Evaluation in the design of health information systems: application of approaches emerging from usability engineering. Computers in Biology and Medicine 2002 May;32(3):141-149. [doi: 10.1016/s0010-4825(02)00011-2]
34. Lee H, Choi B. Knowledge management enablers, processes, and organizational performance: an integrative view and empirical examination. J of Mgmnt Info Sys 2014 Dec 23;20(1):179-228. [doi: 10.1080/07421222.2003.11045756]
35. Harvey G, Kitson A. PARIHS revisited: from heuristic to integrated framework for the successful implementation of knowledge into practice. Implement Sci 2016 Mar 10;11(1):33 [FREE Full text] [doi: 10.1186/s13012-016-0398-2] [Medline: 27013464]
36. Samy GN, Ahmad R, Ismail Z. Security threats categories in healthcare information systems. Health Informatics J 2010 Sep;16(3):201-209 [FREE Full text] [doi: 10.1177/1460458210377468] [Medline: 20889850]
37. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
38. Chan WS, Leung AY. Use of social network sites for communication among health professionals: systematic review. J Med Internet Res 2018 Mar 28;20(3):117 [FREE Full text] [doi: 10.2196/jmir.8382] [Medline: 29592845]
39. Mulgund P, Sharman R, Anand P, Shekhar S, Karadi P. Data quality issues with physician-rating websites: systematic review. J Med Internet Res 2020 Sep 28;22(9):15916 [FREE Full text] [doi: 10.2196/15916] [Medline: 32986000]
40. McLennan S, Strech D, Reimann S. Developments in the frequency of ratings and evaluation tendencies: a review of german physician rating websites. J Med Internet Res 2017 Aug 25;19(8):299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
41. Galletta A. Mastering the semi-structured interview and beyond: from research design to analysis and publication. In: New York University Scholarship Online. Oxfordshire, United Kingdom: Oxford University Press; 2013.
42. Cheung K, van DVW, Bouvy ML, Wensing M, van DBPMLA, de SPAGM. Classification of medication incidents associated with information technology. J Am Med Inform Assoc 2014 Feb;21(e1):63-70 [FREE Full text] [doi: 10.1136/amiajnl-2013-001818] [Medline: 24064444]
43. Kushniruk AW, Triola MM, Borycki EM, Stein B, Kannry JL. Technology induced error and usability: the relationship between usability problems and prescription errors when using a handheld application. Int J Med Inform 2005 Aug;74(7-8):519-526. [doi: 10.1016/j.ijmedinf.2005.01.003] [Medline: 16043081]
44. Leyden J. Medical device vuln allows hackers to falsify patients' vitals. McAfee: Patient monitoring systems open to hack attacks. 2018 Aug 14. URL: https://www.theregister.co.uk/2018/08/14/patient_monitor_hack/ [accessed 2020-08-11]
45. Magrabi, Baker M, Sinha I, Ong MS, Harrison S, Kidd MR, et al. Clinical safety of England's national programme for IT: a retrospective analysis of all reported safety events 2005 to 2011. Int J Med Inform 2015 Mar;84(3):198-206 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.12.003] [Medline: 25617015]
46. Myers RB, Jones SL, Sittig DF. Review of Reported Clinical Information System Adverse Events in US Food and Drug Administration Databases. Appl Clin Inform 2011;2(1):63-74 [FREE Full text] [doi: 10.4338/ACI-2010-11-RA-0064] [Medline: 21938265]
47. Williams P, Woodward A. Cybersecurity vulnerabilities in medical devices: a complex environment and multifaceted problem. MDER 2015 Jul:305 [FREE Full text] [doi: 10.2147/mder.s50048]
48. Ratanawongsa N, Matta GY, Bohsali FB, Chisolm MS. Reducing misses and near misses related to multitasking on the electronic health record: observational study and qualitative analysis. JMIR Hum Factors 2018 Feb 06;5(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.9371] [Medline: 29410388]
49. Ammenwerth E, Schnell-Inderst P, Machan C, Siebert U. The effect of electronic prescribing on medication errors and adverse drug events: a systematic review. J Am Med Inform Assoc 2008;15(5):585-600 [FREE Full text] [doi: 10.1197/jamia.M2667] [Medline: 18579832]
50. Ash J, Sittig DF, Campbell EM, Guappone KP, Dykstra RH. Some unintended consequences of clinical decision support systems. AMIA Annu Symp Proc 2007 Oct 11:26-30 [FREE Full text] [Medline: 18693791]

51. Ash J, Sittig DF, Poon EG, Guappone K, Campbell E, Dykstra RH. The extent and importance of unintended consequences related to computerized provider order entry. J Am Med Inform Assoc 2007;14(4):415-423 [FREE Full text] [doi: 10.1197/jamia.M2373] [Medline: 17460127]
52. Bates D, Cohen M, Leape LL, Overhage JM, Shabot MM, Sheridan T. Reducing the frequency of errors in medicine using information technology. J Am Med Inform Assoc 2001;8(4):299-308 [FREE Full text] [doi: 10.1136/jamia.2001.0080299] [Medline: 11418536]
53. Bates DW. The costs of adverse drug events in hospitalized patients. JAMA 1997 Jan 22;277(4):307 [FREE Full text] [doi: 10.1001/jama.1997.03540280045032]
54. Borycki E, Dexheimer JW, Cossio C, Gong Y, Jensen S, Kaipio J, et al. Methods for addressing technology-induced errors: the current state. Yearb Med Inform 2018 Mar 06;25(01):30-40 [FREE Full text] [doi: 10.15265/iy-2016-029]
55. Campbell E, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. J of the Amer Med Info Asso 2006 Sep 01;13(5):547-556 [FREE Full text] [doi: 10.1197/jamia.m2042]
56. Clarke A, Adamson J, Watt I, Sheard L, Cairns P, Wright J. The impact of electronic records on patient safety: a qualitative study. BMC Med Inform Decis Mak 2016 Jun 04;16:62 [FREE Full text] [doi: 10.1186/s12911-016-0299-y] [Medline: 27260193]
57. Coiera E, Ash J, Berg M. The unintended consequences of health information technology revisited. Yearb Med Inform 2018 Mar 06;25(01):163-169 [FREE Full text] [doi: 10.15265/iy-2016-014]
58. Colpaert K, Claus B, Somers A, Vandewoude K, Robays H, Decruyenaere J. Impact of computerized physician order entry on medication prescription errors in the intensive care unit: a controlled cross-sectional trial. Crit Care 2006 Feb;10(1):R21 [FREE Full text] [doi: 10.1186/cc3983] [Medline: 16469126]
59. Cullen D, Sweitzer BJ, Bates DW, Burdick E, Edmondson A, Leape LL. Preventable adverse drug events in hospitalized patients: a comparative study of intensive care and general care units. Crit Care Med 1997 Aug;25(8):1289-1297 [FREE Full text] [doi: 10.1097/00003246-199708000-00014] [Medline: 9267940]
60. Harrison A, Siwani R, Pickering BW, Herasevich V. Clinical impact of intraoperative electronic health record downtime on surgical patients. J Am Med Inform Assoc 2019 Oct 01;26(10):928-933 [FREE Full text] [doi: 10.1093/jamia/ocz029] [Medline: 30946466]
61. Nebeker JR, Hoffman JM, Weir CR, Bennett CL, Hurdle JF. High rates of adverse drug events in a highly computerized hospital. Arch Intern Med 2005 May 23;165(10):1111-1116 [FREE Full text] [doi: 10.1001/archinte.165.10.1111] [Medline: 15911723]
62. Karsh B, Weinger MB, Abbott PA, Wears RL. Health information technology: fallacies and sober realities. J Am Med Inform Assoc 2010;17(6):617-623 [FREE Full text] [doi: 10.1136/jamia.2010.005637] [Medline: 20962121]
63. Kaushal R, Shojania KG, Bates DW. Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Arch Intern Med 2003 Jun 23;163(12):1409-1416 [FREE Full text] [doi: 10.1001/archinte.163.12.1409] [Medline: 12824090]
64. Koppel R, Metlay JP, Cohen A, Abaluck B, Localio AR, Kimmel SE, et al. Role of computerized physician order entry systems in facilitating medication errors. JAMA 2005 Mar 09;293(10):1197-1203 [FREE Full text] [doi: 10.1001/jama.293.10.1197] [Medline: 15755942]
65. Kumar P, Lee HJ. Security issues in healthcare applications using wireless medical sensor networks: a survey. Sensors (Basel) 2012;12(1):55-91 [FREE Full text] [doi: 10.3390/s120100055] [Medline: 22368458]
66. La Pietra L, Calligaris L, Molendini L, Quattrin R, Brusaferro S. Medical errors and clinical risk management: state of the art. Acta Otorhinolaryngol Ital 2005 Dec;25(6):339-346 [FREE Full text] [Medline: 16749601]
67. Lapetina E, Armstrong EM. Preventing errors in the outpatient setting: a tale of three states. Health Aff (Millwood) 2002;21(4):26-39 [FREE Full text] [doi: 10.1377/hlthaff.21.4.26] [Medline: 12117139]
68. Magrabi F, Ammenwerth E, Hyppönen H, de Keizer N, Nykänen P, Rigby M, et al. Improving evaluation to address the unintended consequences of health information technology. Yearb Med Inform 2018 Mar 06;25(01):61-69 [FREE Full text] [doi: 10.15265/iy-2016-013]
69. Martin G, Ghafur S, Cingolani I, Symons J, King D, Arora S, et al. The effects and preventability of 2627 patient safety incidents related to health information technology failures: a retrospective analysis of 10 years of incident reporting in England and Wales. The Lancet Digital Health 2019 Jul;1(3):127-135 [FREE Full text] [doi: 10.1016/s2589-7500(19)30057-3]
70. Metzger J, Welebob E, Bates DW, Lipsitz S, Classen DC. Mixed results in the safety performance of computerized physician order entry. Health Aff (Millwood) 2010 Apr;29(4):655-663 [FREE Full text] [doi: 10.1377/hlthaff.2010.0160] [Medline: 20368595]
71. Pontefract S, Hodson J, Slee A, Shah S, Girling AJ, Williams R, et al. Impact of a commercial order entry system on prescribing errors amenable to computerised decision support in the hospital setting: a prospective pre-post study. BMJ Qual Saf 2018 Mar 23;27(9):725-736 [FREE Full text] [doi: 10.1136/bmjqs-2017-007135]
72. Schiff G. Medical error: a 60-year-old man with delayed care for a renal mass. JAMA 2011 May 11;305(18):1890-1898 [FREE Full text] [doi: 10.1001/jama.2011.496] [Medline: 21486963]

73. Turner P, Kushniruk A, Nohr C. Are we there yet? Human factors knowledge and health information technology - the challenges of implementation and impact. Yearb Med Inform 2017 Aug;26(1):84-91 [FREE Full text] [doi: 10.15265/IY-2017-014] [Medline: 29063542]
74. Vélez-Díaz-Pallarés M, Álvarez Díaz AM, Gramage Caro T, Vicente Oliveros N, Delgado-Silveira E, Muñoz García M, et al. Technology-induced errors associated with computerized provider order entry software for older patients. Int J Clin Pharm 2017 Aug;39(4):729-742 [FREE Full text] [doi: 10.1007/s11096-017-0474-y] [Medline: 28540465]
75. Wang Y, Coiera E, Gallego B, Concha OP, Ong MS, Tsafnat G, et al. Measuring the effects of computer downtime on hospital pathology processes. J Biomed Inform 2016 Feb;59:308-315 [FREE Full text] [doi: 10.1016/j.jbi.2015.12.016] [Medline: 26732996]
76. Zheng K, Abraham J, Novak L, Reynolds T, Gettinger A. A survey of the literature on unintended consequences associated with health information technology: 2014–2015. Yearb Med Inform 2018 Mar 06;25(01):13-29 [FREE Full text] [doi: 10.15265/iy-2016-036]
77. Flick U. An introduction to qualitative research. University of Michigan 2002:1-310 [FREE Full text]
78. Gonzalez M. The coding manual for qualitative research: a review. The Qualitative Report 2016;21(8):1546 [FREE Full text]

Abbreviations

HIT: health information technology
ICD: International Classification of Diseases
IOT: Internet of Things
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UI: user interface

Edited by B Price; submitted 29.06.20; peer-reviewed by H Hao, O Ben Assuli, A Gupta, E Kasaeyan Naeini, D Lyell; comments to author 19.08.20; revised version received 06.11.20; accepted 17.12.20; published 05.02.21.

Please cite as:
Ndabu T, Mulgund P, Sharman R, Singh R
Perceptual Gaps Between Clinicians and Technologists on Health Information Technology-Related Errors in Hospitals: Observational Study
JMIR Hum Factors 2021;8(1):e21884
URL: http://humanfactors.jmir.org/2021/1/e21884/
doi:10.2196/21884
PMID:33544089

Original Paper

Applying Website Rankings to Digital Health Centers in the United States to Assess Public Engagement: Website Usability Study

Joshua David Calvano[1], BS; Edwin Lauritz Fundingsland Jr[1], BA; Deborah Lai[2], BSEcon; Sara Silacci[3], BS; Ali S Raja[4], MPH, MBA, MD; Shuhan He[3,4], MD

[1]Department of Research, Rocky Vista University College of Osteopathic Medicine, Parker, CO, United States

[2]Division of Psychology and Language Sciences, University College London, Bloomsbury, United Kingdom

[3]Center for Innovation in Digital HealthCare, Massachusetts General Hospital, Boston, MA, United States

[4]Department of Emergency Medicine, Massachusetts General Hospital, Boston, MA, United States

Corresponding Author:
Joshua David Calvano, BS
Department of Research
Rocky Vista University College of Osteopathic Medicine
8401 S Chambers Rd
Parker, CO, 80134
United States
Phone: 1 3033732008
Email: joshua.calvano@rvu.edu

Abstract

Background: As the public increasingly uses the internet to search for resources and information regarding health and medicine, it is important that health care organizations provide adequate web resources. Website usability refers to the ease of user experience on a website. In this study, we conducted usability analyses on digital health center websites.

Objective: The primary aims of this study were to (1) replicate a preexisting usability scoring methodology for digital health centers; (2) apply and test this replicated usability scoring methodology on a sample set of digital health center websites; and (3) derive recommendations from the results on potential areas of improvements for our sample of digital health center websites.

Methods: Website usability testing was conducted from March 1, 2020, to March 15, 2020. We replicated a methodology and scoring system from previous literature and applied them to digital health center websites. Our sample included 67 digital health centers that were affiliated with US universities or hospital systems. Usability was split into the following four broad categories: accessibility, marketing, content quality, and technology. Usability tools were used to score websites in each of the four categories. The composite of the key factors of each category was used to generate a general usability and overall usability score for each website.

Results: The category with the highest average score (6.3) was content quality. The content quality score also had the highest SD (2.18) and an SE of 0.27. The lowest performing category was technology, which had an average score of 0.9. The technology score also had the smallest SD (0.07) and an SE of 0.01.

Conclusions: Our data suggest that content quality, on average, was the highest scoring variable among digital health center websites. As content is crucial to digital health knowledge, it is justified that digital health centers invest more resources into creating quality content. The overall lowest scoring variable was technology. Potential reasons for this finding include designated funding for servers, a lack of regulatory frameworks for social media presence and liability, and infrequent website audits. An easy approach for improving this variable is increasing website speed. Accessibility is another area that organizations can potentially improve. We recommend that these organizations perform periodic audits of their web presence with usability tools.

(JMIR Hum Factors 2021;8(1):e20721) doi:10.2196/20721

KEYWORDS

website usability; digital health; health care website; usability testing; web interventions; digital health care; web crawler

Introduction

Background

A hospital's or digital health care center's website is often these organizations' first point of contact with the public; therefore, websites are crucial in first impressions [1,2]. They have the potential to be an important part of the first step in improving patient satisfaction and attracting new patients [3]. In a time when information is expected to be readily available, health care organizations use their websites as key tools for both patient communication and education [4-6]. Patients expect to find current and reliable information on websites that are easily accessible in order to make health-related decisions [7]. As many health-related sources are available (eg, WebMD), health care organizations are aiming to improve their internet presence so that they can better communicate with and market to potential customers [3].

Website Usability

Improving website usability is a noteworthy approach that medical organizations can use to improve their internet presence, attract and retain more users, and disseminate accurate and reliable information to a larger audience. Usability goes beyond surface-level design; it broadly refers [8] to a product's user experience, which includes aspects such as the ease of navigation or user-encountered problems within a website [9]. It addresses the question of how easy or pleasing a website is to use, which are factors that can influence the number of users that engage with a website. Usability also addresses users' level of engagement and a website's ability to achieve other objectives. When users are not able to easily access and use a website, they are unlikely to continue using it as an information source. Alternatively, improved usability can enhance the reach of a website. It is for this reason that websites are facing the increasing need to conform to user expectations, desires, and requirements [10,11]. Various industries have established standardized guidelines for accessibility, content, marketing, and technology to improve website usability [12-14].

Usability Studies for Digital Health Centers

Studies have sought to apply usability analyses to e-commerce, e-governments, mobile news apps, and library websites [15-18]. In health care, other studies have analyzed the usability of hospital, children's hospital, and cancer center websites [3,19,20]. However, to our knowledge, no usability studies have been conducted for digital health centers in the United States. Digital health centers combine innovation-driven health care research with digital technology. Digital health technologies are emerging tools that have the potential to improve patient-centered health care by improving care quality and reducing health care costs [21]. Given digital health centers' focus on digital technologies (eg, technologies for improving web presence), there is an opportunity to better understand how digital health centers are adapting to technologies that use their web presence. Specifically, we believe that it is distinctly important for these organizations to create websites with high usability to not only improve user experience but also present themselves as leaders in innovation.

Objectives

The primary aims of this study were (1) to replicate a preexisting usability scoring methodology for digital health centers; (2) to apply and test this replicated usability scoring methodology on a sample set of digital health center websites; and (3) to derive recommendations from the results on potential areas of improvement for our sample of digital health center websites.

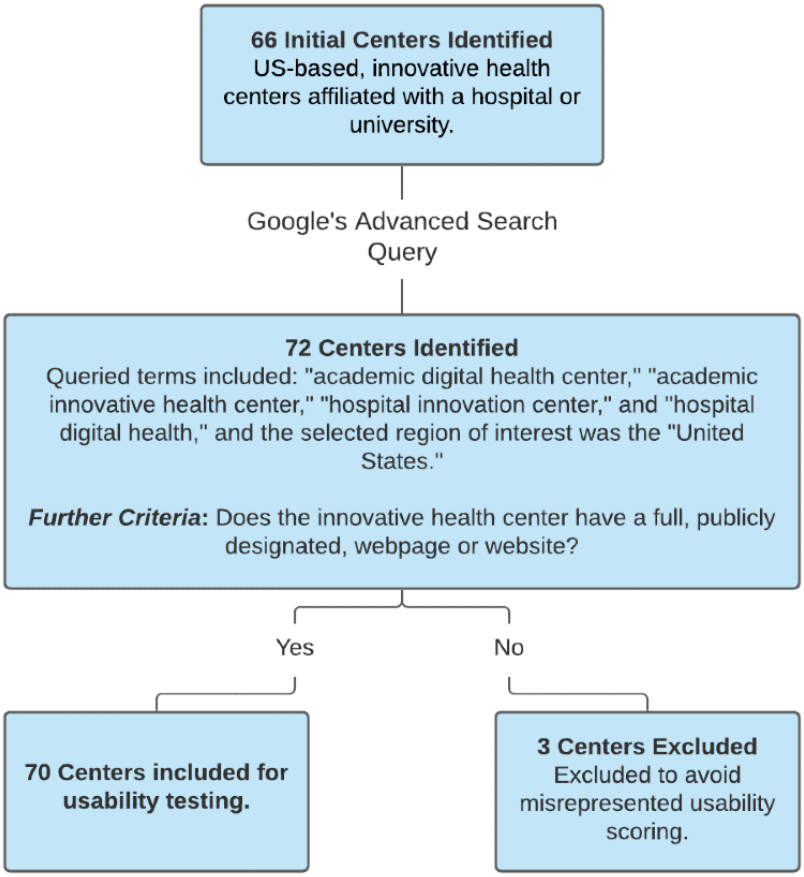
Methods

Sample Selection

Our focus was on digital health centers that were affiliated with US universities or hospital systems. Indexing the websites of all digital health centers, such as medical companies, was not within the scope of our study.

The original sample set was derived from Becker's Hospital Review and consisted of a total of 66 digital health centers [22]. We augmented this sample set by including 8 additional digital health centers that were found with Google's Advanced Search query builder, which increased the total number of digital health centers in our sample to 72. The terms and phrases searched included "academic digital health center," "academic innovative health center," "hospital innovation center," and "hospital digital health," and the selected region of interest was the United States. We excluded three digital health centers that did not have a designated digital health center website. Our final sample set consisted of 70 digital health centers (Figure 1).

Figure 1. Sample selection criteria for digital health center websites.



Overview

Website usability tests were conducted from March 1, 2020, to March 15, 2020. The methodology used in this study was replicated from previously published health care usability literature [3,19]. We chose to evaluate the same factors as those in the previous studies. However, we modified the definitions for clarity and reproducibility by using our selected assessment tools (Multimedia Appendix 1). We used the same weighted percentages as those from the previous studies and applied specific formulas to these calculated percentages to create a relative scale for comparing usability scores.

In alignment with the replicated methodology, websites were assessed with four scales for the following categories: (1) accessibility, which refers to the ability of people with low levels of computer literacy to access and navigate hospital's web presence; (2) marketing, which refers to the ability to be found through search engines and the relevance of descriptions to the links provided; (3) content quality, which refers to grammar, the frequency of information updates, material relevancy, and readability; and (4) technology, which refers to download speed, the quality of the programming code, and website infrastructure [3,19]. Each of these categories represent distinct, quantifiable, and actionable areas of usability that digital health centers can improve on to communicate more effectively with their audiences.

Analysis

All websites were analyzed by using a set of established usability tools (Multimedia Appendix 1). The tools were chosen based on their ability to meet the industry standards for evaluating the selected factors and their relative ease of use. The process for using each tool was based on the tool's specific instruction manual. One author of this manuscript supervised the training for a team of 6 student reviewers. Each reviewer underwent the same training for using the suite of analytic tools and performing data entry. The reviewers were then directly observed while they used each tool on three websites, in order to confirm proper usage and reliability. Discrepancies and questions were addressed and answered by the supervising author as they arose. In addition, each tool was used on the same local computer to account for irregularities such as differing internet service providers or computing components, which might affect the consistency and reliability of the results. Factors that can vary from second to second, such as speed, were averaged across two separate tools to provide the most accurate values possible.

We built a database of the top-level URLs that were associated with each website in our data set. This was done by using a web crawler, which is a tool that processes URLs and creates topographical maps of a website and all of its subpages. For instance, a top-level domain that corresponds to a website's home page may be associated with the URL www.healthcare.org. A subpage of this center's website might be a page about the team members, which might be associated with the URL www.healthcare.org/team. There may be other

subpages for specific topics, such as the emergency medicine department and the pediatric department. Once the web crawler creates a topographical map of a website, that website can then be analyzed for page errors, the amount of page content, metadata (ie, titles, keywords, and descriptions), or other preprogrammed factors [23].

Websites in our data set received a final score for the following four categories: accessibility, content, marketing, and technology. Per the replicated methodology, the composite of select key factors across each of the four categories was used to generate a fifth general usability score for each website. A weighted aggregate of these five scores was used as the sixth and final score for the final ranking system.

In the following sections, we describe each of the categories that we evaluated, the development of the rating scale for each of these categories, and the importance of each category's contributions.

Accessibility

The accessibility rating indicates a website's appeal to a broad audience of people with varying literacy levels, technical aptitudes, and disabilities. This category involves factors such as a web page's meta-description, readability, and the overall layout of the website. A meta-description is the "snippet" page summary that appears in search results when using a search engine. Another factor is functionality, which encompasses elements like actionable buttons that send users to parts of a website and content that is understandable to people with a wide range of education levels or reading abilities. For instance, it has been reported that an estimated 43% of American adults have basic or below basic literacy levels [24]. Accessibility ratings can also be used to evaluate the usability of assistive technologies, such as screen readers and magnifiers for a given website [25]. For our study, we used the Flesch-Kincaid Reading Ease and Gunning Fog Index algorithmic readability scales to rank a website's reading difficulty and approximate the grade level required to understand each website's content.

Content Quality

The content quality rating is used to assess the content on a website. This can include the relevancy of written information to a particular point in time and a specific topic, generated metadata, and the use of a website's multimedia elements. For instance, a website that is dedicated to supplying information on current closed-loop insulin pumps for people with diabetes may be evaluated on its ability to provide relevant, fact-driven answers to people who seek such information (eg, relative costs, ease of use, etc). In content quality analysis, the multimedia elements on a website may be evaluated for their quality (eg, resolution) and their ability to support the website's content (ie, available metadata functions). Content quality analysis also involves the assessment of written text (ie, the evaluation of grammar and spelling).

Marketing

The marketing rating indicates the discoverability of a website. This rating puts particular emphasis on search engine results pages (SERPs), which refer to websites that are suggested to users when they search for information via a web-based search engine, such as Google. Higher placement in search results can lead to greater visibility, and SERPs are considered by some as one of the most important elements of digital marketing. The field of search engine optimization (SEO) involves optimizing a website to achieve higher placement in SERPs, and effectively implementing SEO methods may allow health care organizations to uphold their corporate image as industry leaders [26]. Technical SEO auditing was beyond the scope of this study.

Technology

The technology rating indicates the technical functionality of a website as opposed to its content quality; it indicates the quality of a website's technology, technological design, and performance. The technology rating encompasses various aspects, including front-end design, user experience, back-end coding infrastructure, and server management. The front end is what users view when browsing a website. Front-end design involves analyzing HTML elements to ensure that a user is provided with an easily navigable layout and that the website is scalable across devices (ie, computers, mobile phones, and tablets). The back end refers to the programming code upon which the website runs. This code and other website components, such as databases, are stored on servers, which functionally allow people to view websites from their own devices. The servers also affect the speed of the website (eg, how quickly it loads for users), which can play a crucial role in gaining and maintaining users and followers. For instance, a previous study conducted by Google [27] showed that a website that takes longer than 3 seconds to load on a mobile device will lose approximately 53% of its users. Furthermore, the study revealed that the average mobile website speed is upwards of 18 seconds [27].

General Usability

The general usability rating was based on a composite of select key factors from the prior four categories. The concept of general usability aims to answer the question "how good is my website?" This metric may serve as a starting point for health care organizations to perform an initial audit of their website and identify areas of improvement.

Overall Usability

An overall usability rank order calculation was performed to comprehensively evaluate all major and minor factors across the five aforementioned categories. Afterward, we assigned weighted percentages to all factors to create an all-inclusive usability ranking system.

Results

Technical issues prevented the web crawler from running on three websites. This was possibly due to the fact that no index restrictions were set up by the website administrators. We assigned scores to the remaining (N=67) digital health centers.

The subcategory with the highest average score (6.3) was content quality. The content quality score also had the highest SD (2.18) and an SE of 0.27. Accessibility was the second highest scoring subcategory, which had an average score of 2.2.

The accessibility score had an SD of 0.51 and an SE of 0.06. Of the four subcategories, marketing had the third highest average score (1.5). The marketing score had an SD of 0.40 and an SE of 0.05. The lowest performing subcategory was technology, which had an average score of 0.9. The technology score also had the smallest SD (0.07) and an SE of 0.01. The summary statistics for all five categories are presented in Table 1.

Table 1. Digital health center website summary statistics from the usability analysis.

Category	Score, mean (SE; SD)	Score, range
Accessibility	2.2 (0.06; 0.51)	0.9-3.3
Content quality	6.3 (0.27; 2.18)	1.1-10.7
Marketing	1.5 (0.05; 0.40)	0.6-2.4
Technology	0.9 (0.01; 0.07)	0.7-1.0
General usability	1.5 (0.04; 0.33)	0.8-2.2

The overall rankings for the 67 assessed domains across all categories are reported in Multimedia Appendix 2. The highest scoring centers across all five usability ranking categories were as follows: (1) Sutter Health Design and Innovation (accessibility score=3.3); (2) Sutter Health Design and Innovation (content quality score=10.7); (3) Mayo Clinic Center for Innovation (marketing score=2.4); (4) University of Texas Southwestern Office for Technology Development (technology score=1); and (5) Sutter Health Design and Innovation (general usability score=2.2). In terms of overall usability, the top performing website was that of Sutter Health Design and Innovation (overall usability score=3).

Discussion

Comparison With Prior Work

Emerging technologies in the field of digital health are rapidly changing the aspects of health care by making them more patient centered, improving care quality, and decreasing health care costs [21]. The increasing importance of digital health has made it an appropriate field for website usability research.

Our study involved methods that were replicated from previous studies. This allowed us to assess similar trends across various health care website dimensions [3,19]. As with previous studies, the overall scores in our study were highest for the content quality category. This finding could reflect the importance of information to the health care industry and indicate that health centers should invest most heavily in content quality when creating their websites.

Another finding that was consistent with prior research was that the overall lowest ranking category was technology [3,20]. This may be due to a lack of investment in digital technology by the health care industry (eg, investments in server capacity, social media, and website audits). One approach for immediately improving technology is improving website speed. This is largely accomplished through modifying back-end web server settings and minimizing the number of conflicting technologies that run on a website.

A study that evaluated children's hospitals found accessibility to be the lowest ranking category, which was not the case in our study. However, our accessibility scores were lower than originally anticipated [19]. With regard to health care, we believe that accessibility should be paramount. Health industry leaders should put more effort into ensuring that all domains remain functional and accessible to everyone, so that the quality content on these websites can reach the appropriate users [25].

Limitations

This study has several limitations. It is common for large organizations to have specific subpages that are dedicated to digital health. For example, one website had an estimated domain age of 33 years when, in reality, the associated innovation health center was aged less than 5 years. However, structuring websites in this manner may provide digital health centers with a competitive advantage, as this method would improve their rankings. This would result in an increase in the number of people who view their website.

Additional limitations and difficulties were found in the assessment of social media websites. Not all social media accounts were directly accessible from these websites. As such, it was difficult to find certain social media accounts through Twitter's and Facebook's respective search engines. Oftentimes, the digital health centers' profiles were distant from the top result.

Assessments of a website's speed can vary depending on the time of day or the day of data collection. This could be due to changes in the website's servers, internet connectivity, or computer hardware. To minimize sampling bias, the same computer and the same network were used for all of our tools.

Another limitation was that all information was collected over the course of 2 weeks. As such, several measures may have changed since the initial evaluation.

Conclusion

With digital health emerging as a leading field in terms of innovation in health care, it is important that digital health centers are able to effectively connect with the public by using their websites. In this study, we conducted an analysis of the overall need for improving the usability of digital health centers' websites. The average general usability score was 1.5. This shows the necessity of improving the usability aspects of websites. Digital health centers may benefit from taking steps to improve the various components of their websites in order to reach their audiences. A suggested step for these organizations is to perform periodic usability audits of their websites to

identify areas for improvement. Several of these institutions have considerable room for improvement in terms of their overall web presence. We have identified approaches that these organizations can use to increase their websites' usability, such as improving website speed and social media access. These approaches could potentially improve their websites' reach.

Acknowledgments

We would like to thank the following individuals from the Rocky Vista University College of Medicine for their contributions in data entry and ongoing support: Brandon Bealer, Kameron Black, Joseph Fike, Andrew Glerum, and Ryan Henschell.

Conflicts of Interest

SS and SH helped to create the website for the Massachusetts General Hospital Center for Innovation in Digital HealthCare. The website was excluded from this study.

Multimedia Appendix 1
Defined usability factors with their associated percentage weights, assessment tools, impacts, and formulas.
[PDF File (Adobe PDF File), 36 KB - humanfactors_v8i1e20721_app1.pdf]

Multimedia Appendix 2
Digital health center websites and category scores.
[PDF File (Adobe PDF File), 91 KB - humanfactors_v8i1e20721_app2.pdf]

References

1. Alpay LL, Overberg RI, Zwetsloot-Schonk B. Empowering citizens in assessing health related websites: a driving factor for healthcare governance. International Journal of Healthcare Technology and Management 2007;8(1/2):141. [doi: 10.1504/ijhtm.2007.012103]
2. Revere L, Robinson L. How healthcare organizations use the Internet to market quality achievements. J Healthc Manag 2010;55(1):39-49; discussion 49. [Medline: 20210072]
3. Huerta TR, Hefner JL, Ford EW, McAlearney AS, Menachemi N. Hospital website rankings in the United States: expanding benchmarks and standards for effective consumer engagement. J Med Internet Res 2014 Feb 25;16(2):e64 [FREE Full text] [doi: 10.2196/jmir.3054] [Medline: 24568892]
4. Coile RC. E-health: reinventing healthcare in the information age. J Healthc Manag 2000;45(3):206-210. [Medline: 11066968]
5. Randeree E, Rao HR. E-health and assurance: curing hospital websites. Int J Electron Healthc 2004;1(1):33-46. [doi: 10.1504/IJEH.2004.004653] [Medline: 18048202]
6. Wen K, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010 Dec 18;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
7. Reid P, Borycki EM. Emergence of a new consumer health informatics framework: introducing the healthcare organization. Stud Health Technol Inform 2011;164:353-357. [Medline: 21335736]
8. Griffiths KM, Christensen H. Website quality indicators for consumers. J Med Internet Res 2005 Nov 15;7(5):e55 [FREE Full text] [doi: 10.2196/jmir.7.5.e55] [Medline: 16403719]
9. Otis M, Zhu J, Mustafa-Kutana SN, Bernier AV, Shum JM, Dupre JAS, et al. Testing usability and feasibility of a mobile educator tool for pediatric diabetes self-management: Mixed methods pilot study. JMIR Form Res 2020 May 01;4(5):e16262 [FREE Full text] [doi: 10.2196/16262] [Medline: 32356773]
10. Mittler JN, Volmar KM, Shaw BW, Christianson JB, Scanlon DP. Using websites to engage consumers in managing their health and healthcare. Am J Manag Care 2012 Sep;18(6 Suppl):s177-s184. [Medline: 23286713]
11. Liang C, Chen H. A study of the impacts of website quality on customer relationship performance. Total Quality Management & Business Excellence 2009 Sep;20(9):971-988. [doi: 10.1080/14783360903181784]
12. Oermann MH, Lowery NF, Thornley J. Evaluation of Web sites on management of pain in children. Pain Manag Nurs 2003 Sep;4(3):99-105. [doi: 10.1016/s1524-9042(03)00029-8] [Medline: 14566707]
13. Oermann MH, Lesley ML, VanderWal JS. Using Web sites on quality health care for teaching consumers in public libraries. Qual Manag Health Care 2005;14(3):188-195. [doi: 10.1097/00019514-200507000-00008] [Medline: 16027597]
14. Oermann MH, McInerney SM. An evaluation of sepsis Web sites for patient and family education. Plast Surg Nurs 2007;27(4):192-196. [doi: 10.1097/01.PSN.0000306184.95812.8a] [Medline: 18165726]
15. Kuan H, Bock G, Vathanophas V. Comparing the effects of usability on customer conversion and retention at e-commerce websites. : IEEE Computer Society; 2005 Presented at: Proc Annu Hawaii Int Conf Syst Sci; January 6, 2005; Big Island p. 174 URL: https://ieeexplore.ieee.org/document/1385584 [doi: 10.1109/hicss.2005.155]

16. Huang Z, Benyoucef M. Usability and credibility of e-government websites. Gov Inf Q 2014 Oct;31(4):584-595. [doi: 10.1016/j.giq.2014.07.002]
17. Jeong W, Jung Han H. Usability study on newspaper mobile websites. OCLC Systems & Services 2012 Oct 26;28(4):180-198. [doi: 10.1108/10650751211279120]
18. N VR, Harinarayana NS. An analysis of usability features of library web sites. Annals of Library and Information Studies 2008 Jun;55(2):111-122 [FREE Full text]
19. Huerta TR, Walker DM, Ford EW. An evaluation and ranking of children's hospital websites in the United States. J Med Internet Res 2016 Aug 22;18(8):e228 [FREE Full text] [doi: 10.2196/jmir.5799] [Medline: 27549074]
20. Huerta TR, Walker DM, Ford EW. Cancer center website rankings in the USA: Expanding benchmarks and standards for effective public outreach and education. J Cancer Educ 2017 Jun;32(2):364-373. [doi: 10.1007/s13187-015-0931-z] [Medline: 26472325]
21. DePasse JW, Chen CE, Sawyer A, Jethwani K, Sim I. Academic medical centers as digital health catalysts. Healthc (Amst) 2014 Sep;2(3):173-176. [doi: 10.1016/j.hjdsi.2014.05.006] [Medline: 26250503]
22. Dyrda L. 66 hospitals and health systems with innovation programs. Becker's Hospital Review. URL: https://www.beckershospitalreview.com/lists/66-hospitals-and-health-systems-with-innovation-programs-2018.html [accessed 2020-05-02]
23. Devi RS, Manjula D, Siddharth RK. An efficient approach for web indexing of big data through hyperlinks in web crawling. ScientificWorldJournal 2015;2015:739286 [FREE Full text] [doi: 10.1155/2015/739286] [Medline: 26137592]
24. Kutner M, Greenberg E, Jin Y, Boyle B, Hsu Y, Dunleavy E. Literacy in everyday life: Results from the 2003 National Assessment of Adult Literacy. National Center for Education Statistics. 2007. URL: https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2007480 [accessed 2021-03-25]
25. Ismail A, Kuppusamy KS, Nengroo AS. Multi-tool accessibility assessment of government department websites:a case-study with JKGAD. Disabil Rehabil Assist Technol 2018 Aug;13(6):504-516. [doi: 10.1080/17483107.2017.1344883] [Medline: 28766367]
26. Batraga A, Rutitis D. Corporate identity within the health care industry. Economics and Management 2012 Dec 07;17(4):1551. [doi: 10.5755/j01.em.17.4.3027]
27. New Industry Benchmarks for Mobile Page Speed. Think with Google. URL: https://www.thinkwithgoogle.com/marketing-resources/data-measurement/mobile-page-speed-new-industry-benchmarks/ [accessed 2020-05-01]

Abbreviations

SEO: search engine optimization
SERP: search engine results page

Edited by G Eysenbach; submitted 26.05.20; peer-reviewed by M Nitsch; comments to author 22.07.20; revised version received 02.08.20; accepted 02.03.21; published 29.03.21.

Please cite as:
Calvano JD, Fundingsland Jr EL, Lai D, Silacci S, Raja AS, He S
Applying Website Rankings to Digital Health Centers in the United States to Assess Public Engagement: Website Usability Study
JMIR Hum Factors 2021;8(1):e20721
URL: https://humanfactors.jmir.org/2021/1/e20721
doi:10.2196/20721
PMID:33779564

Original Paper

The Relationship Between Individual Coping and the Need to Have and Seek Health Information Among Older Adults: Exploratory Mixed Methods Study

Sabine Theis[1], Dr-Ing; Katharina Schäfer[1], MA; Dajana Schäfer[1], MSc; Nicole Jochems[2], Univ-Prof, Dr-Ing; Verena Nitsch[1], Univ-Prof, Dr-Ing; Alexander Mertens[1], Dr-Ing, Dr rer medic, Dipl-Inform

[1]Institut für Arbeitswissenschaft, RWTH Aachen University, Aachen, Germany

[2]Institut für Multimediale und Interaktive Systeme, Universität zu Lübeck, Lübeck, Germany

Corresponding Author:
Sabine Theis, Dr-Ing
Institut für Arbeitswissenschaft
RWTH Aachen University
Eilfschornsteinstraße 18
Aachen, 52062
Germany
Phone: 49 241809946
Email: s.theis@iaw.rwth-aachen.de

Abstract

Background: The need to have and seek information shapes the context of computing systems. When it comes to health, individual coping influences human behavior. Therefore, the relationship between individual coping and the need to have and seek health information plays a crucial role in the development of digital health systems.

Objective: This study aims to examine the relationship between individual coping and the need to have and seek health information among older adults.

Methods: Questionnaires and semistructured interviews investigated the health information need (HIN) and health information–seeking behavior (HISB) in relation to the individual coping strategies of 26 older Germans.

Results: The mean age of the interviewed group was 71 years (SD 7). Quantitatively, a trend was found for a negative correlation between the avoidance-oriented coping and HIN (r_s=−0.37895; bias-corrected and accelerated bootstrap 95% BCa CI −0.730 to 0.092; *P*=.05). The qualitative results supported this finding. For some participants, information and exchange was part of dealing with their health situation, whereas others wanted to learn as little as possible to avoid a decline in their health status. The older adults acquired, collected, and exchanged paper-based health data to augment clinical information sources and support information exchange with professionals.

Conclusions: Individual coping strategies are relevant for the design of digital health systems. They can support older adults in coping with their health situation, although it remains unclear how systems must be designed for people with an avoidance coping strategy to achieve the same acceptance.

(JMIR Hum Factors 2021;8(1):e15858) doi:10.2196/15858

KEYWORDS

health; information science; systems analysis; eHealth; engineering; gerontology; information technology; mobile phone

Introduction

Background

Due to demographic changes and the underlying aging society, the number of people in need of help and care increases. At the same time, however, the number of nursing staff decreases and a gap emerges that can hardly be closed by the care provided by family members alone [1]. To address this problem, experts place great hope in health digitalization. Digital health systems offer an opportunity to support and maintain the independence and self-responsibility of older people; they enable professional health services to be made more effective and family members to be relieved [2]. Therefore, an analysis of the use context is necessary and is the subject of this study.

Cognition, knowledge, and personal experiences of the user [3,4]; the working environment; and the user's task are a few of the standard variables considered when investigating the context of information systems. However, this context can also be described by the target group's health information need (HIN) and health information–seeking behavior (HISB) [5-9]. HISB denotes the search for health information resulting from a perceived HIN to reach a certain goal [10]. When seeking information, a user applies different sources of information that might be analogous, such as a print medium (eg, the newspaper) or a digital medium such as the internet or a smartphone app [11,12]. In fact, the HIN of older adults showed a relationship with the use of health information seeking [13]. Here, it appeared that older adults, who require more information about their health, engage more with mobile devices such as smartwatches or mobile health apps installed on tablet PCs. However, the influences on HIN seem to be manifold.

In the health care context, where an illness often relates to a stressful situation, individual coping strategies can have an influence on patients' HISB [14-16]. In particular, the model of information-seeking behavior by Wilsons and Walsh [17] illustrates that actions or a series of actions taken to approach an unpleasant or stressful situation include or are related to HISB. For example, people who want to avoid coping with their illness feel more overwhelmed by illness-related information and less if provided by an information presentation that fits their coping strategy. Not surprisingly, van Zuuren and Wolfs [18] found that HISB is highly related to task- and problem-oriented coping strategies and that some people even perceive the information itself as a threat similar to the illness. Lower socioeconomic status, poor health, low media attentiveness, and high affective components of information seeking were associated with overload. The strongest predictors were education level and cognitive aspects of information seeking, which indicates that health information literacy skills strongly predict the overload [19].

Research Questions

The incidence of disease increases during the course of life. In older adults, illnesses occur more frequently from the age of 50 years. Older adults thus represent an important audience for digital health systems. Although previous work described their information needs and behavior quantitatively, qualitative descriptions and relationships with individual coping are lacking. As individual coping has a particularly strong influence on a person's behavior in the context of an illness, the following research questions (RQs) investigate the information needs (RQ 1) and information seeking (RQ 2) of older adults and their relation to individual coping (RQ 3) in a qualitative and quantitative manner: RQ 1: Which HINs do older adults have?, RQ 2: How do older adults acquire the needed health information?, RQ 3: How does the coping of older adults relate to their HINs and HISB?

Methods

Study Design

To answer the previously mentioned RQs, a mixed methods field study was conducted [20]. Qualitative interviews allow respondents to talk at some depth, choosing their own words to describe their HIN and HISB. Questionnaires then quantitatively measured HIN, HISB, and individual coping (Coping Inventory of Stressful Situations [CISS]).

Participants

The sample consisted of 26 older adults living in the German state of North Rhine-Westphalia. A total of 18 interviewees also answered the questionnaires. Moreover, 8 participants answered the questionnaires only, as they refused to be interviewed directly. A total of 3 interview recordings (ID01, ID11, and ID12) were lost because of recording issues.

A total of 33% (5/15) of the interviewed participants who answered the questionnaire were male and 67% (10/15) were female. The mean age of the interviewed group was 71 years (SD 7). The participants had a rather varied level of education: 14 had completed secondary modern school (*Volksschule/Hauptschule* in German). Five of these had subsequently undergone vocational training (*Berufsausbildung*). A total of 3 participants attended secondary school (*Realschule*) and high school (*Gymnasium*), whereas only 1 participant had a university degree (*Hochschulabschluss*). Participants ID04 and ID05 were a couple and interviewed together.

All 26 participants were born in Germany. Participants were primarily office employees and craftsmen. In total, 8 of the participants who were interviewed and answered the questionnaires had a leadership position, whereas 7 did not have a leadership position. A total of 54% (14/26) were living at home with their partner, wife, or husband. The other 31% (8/26) lived alone at home, and 15% (4/26) lived at retirement homes (Table 1).

Table 1. Demographics (N=26).

ID	Age (years)	Gender	Educational level	Living situation	Interview
01	88	Female	Secondary modern school	At retirement home	Lost data
02	74	Male	High school	At home, with spouse	Completed
03	68	Female	Vocational training	At home, with spouse	Completed
04	80	Male	Secondary modern school	At home, with spouse	Completed
05	78	Female	Secondary modern school	At home, with spouse	Completed
06	82	Female	Secondary school	At home, alone	Completed
07	68	Female	High school	At home, alone	Completed
08	76	Female	Vocational training	At home, alone	Completed
08	77	Female	Secondary school	At home, with spouse	Completed
10	61	Male	Vocational training	At home, with spouse	Completed
11	75	Female	Secondary modern school	At home, with spouse	Lost
12	80	Male	Secondary modern school	At home, with spouse	Lost
13	64	Female	Secondary modern school	At home, alone	Completed
14	64	Female	Secondary modern school	At home, with spouse	Completed
15	66	Male	Secondary modern school	At home, with spouse	Completed
16	65	Male	Vocational training	At home, with spouse	Completed
17	65	Female	Vocational training	At home, with spouse	Completed
18	78	Female	PhD	At home, alone	Completed
19	76	Female	Secondary modern school	At home, alone	Refused
20	79	Female	Secondary modern school	At home, with spouse	Refused
21	72	Female	Secondary modern school	At home, with daughter	Refused
22	68	Female	Secondary modern school	At home, alone	Refused
22	72	Male	High school	At home, alone	Refused
24	83	Female	Secondary school	Retirement home	Refused
25	92	Male	Secondary modern school	Retirement home	Refused
26	79	Female	Secondary modern school	Retirement home	Refused

Procedure

The interviews were conducted and the questionnaires were answered during the interviewers' visit to the participants' homes. During the visit, the inquiry procedure took up to 1.5 hours and started with an introduction, followed by acquiring informed consent and answering demographic questions. Subsequently, a semistructured interview was conducted, which was followed by different structured questionnaires and, finally, by the assessment of individual coping strategies via the CISS.

Interview Guideline

The qualitative interview guideline was based on the study by Warner et al [21] (Multimedia Appendix 1), who investigated nonoccupational information needs using a framework that focused on the essential components of information needs and behaviors—the user, the needs, the sources of information, and the tools and solutions users apply—and on the interaction effects between these variables. Their tool was pretested with data from a cross-sectional and large random sample. It was adapted to the domain of personal health and shortened to the version included in Multimedia Appendix 1.

Initially, participants were queried on their required health information. Subsequently, they were asked about the information they needed during the last week or month regarding their personal health or health in general, vital data, medication prevention, treatment, and additional topics that occurred to the participants. In section 2 of the interview, participants had to rank the mentioned information needs and describe the frequency and time of occurrence of the question or the problem they had and the sources they already used, planned to use, or failed to use. In cases where no information sources were acquired, participants were asked to describe which sources they thought might have the necessary information.

Sociodemographic Questionnaire

With this questionnaire, we queried not only standard but also theoretically relevant parameters such as age, educational achievements, professional background and available

information sources, cultural background, and current living conditions (Multimedia Appendix 1).

CISS

The CISS is a 48-item instrument used to measure 3 basic coping strategies, with 16 items per scale: task-oriented (T), emotion-oriented (E), and avoidance (A) [22,23]. Items are scored on a 5-point Likert scale (from 1=not at all to 5=very much). Higher scores indicate greater use of that particular coping strategy. To exclude the interviewee's fatigue as an external factor, we decided to apply the *paper-and-pencil* MHS QuikScore with 21 items.

Information Need Questionnaire

In addition to the open, semistructured interview, the need for information and the behavior were queried using a specially created, theory-based questionnaire [17]. The need for information is implicitly taken into account by the question of satisfaction with the available health information, which the participants were able to answer using a 5-point Likert scale (1=applicable, 2=rather applicable, 3=partially, 4=rather not applicable, and 5=not applicable). The corresponding questionnaire can be found in Multimedia Appendix 1.

Qualitative Data Analysis

Theoretical thematic analysis inspired by the 6-step recursive process by Braun and Clarke was used to analyze the qualitative interview data. The main advantages of thematic analysis lie in its flexibility, usefulness, and easy access to researchers who are new to qualitative research [24]. Transcripts were analyzed thematically. Thematic analysis is characterized by an essentialist, analyst-driven, and semantic approach, which means that the coding process was done in relation to the RQs, preresearched concepts of HIN and HISB, and, thus, with regard to particular areas of interest. Progression from a semantic level to a level of interpretation gave rise to broader meanings and implications. With respect to the HIN and HISB topics, we systematically coded with the help of *Dedoose* software (University of California, Los Angeles). There were no qualitative questions on individual coping during the interviews. Individual coping was quantitatively measured using a questionnaire (CISS). The relationship between HIN and HISB and coping was analyzed by quantitatively grouping participants into the coping groups, as described in the following section, and then qualitatively describing the HIN and HISB groups.

Quantitative Analysis

To investigate the influence of individual coping strategies on HIN and HISB of older adults, we built CISS-type clusters based on the 3 CISS dimensions: *task-oriented coping* (T), *emotion-oriented coping* (E), and *avoidance-oriented coping* (A). When a participant's score of each of the 3 was higher than the mean of all participants on the same dimension, this dimension labeled the dimension type. This resulted in 6 CISS types—T, E, A, TE (task-emotion oriented coping), TA (task-avoidance oriented coping), and TEA (task-emotion-avoidance oriented coping)—based on which we compared the questionnaire results on information need and seeking behavior (Wilson Questionnaire given in Multimedia Appendix 1).

The descriptive analysis of the questionnaire data was carried out using Dedoose software (University of California, Los Angeles). The statistical software SPSS (IBM Corp) version 24 was used to calculate the chi-square test results attaining the relation of HIN, HISB, and CISS subscales. The CISS groups were compared qualitatively within the framework of a mixed methods analysis. An overview of the qualitative and quantitative measures and the analysis with respect to the independent variables is depicted in Table 2.

Table 2. Overview of the quantitative and qualitative measurement and analysis methods.

Measurement and analysis type	HIN[a]	HISB[b]	Coping×HIN/HISB
Measurement: QUAL[c]	Interview	Interview	N/A[d]
Measurement: QUAN[e]	Questionnaire	Questionnaire	Questionnaire
Analysis: QUAL	Thematic analysis	Thematic analysis	N/A
Analysis: QUAN	Descriptive statistics	Descriptive statistics	Grouping, correlation
Analysis: QUAL+QUAN	N/A	N/A	Thematic analysis of groups

[a]HIN: health information need.

[b]HISB: health information–seeking behavior.

[c]QUAL: qualitative.

[d]N/A: not applicable.

[e]QUAN: quantitative.

Results

HIN Interview Data

In total, 3 groups of information need emerged from the thematic analysis. The participants themselves indicated the level of HIN either directly or indirectly, for example, by naming a lot of information needs. A total of 2 independent qualitative analysts assigned them to the corresponding category. In cases where different assignments were made, the decision was discussed and then a unanimous decision was made. The following groups emerged from the analysis: (1) participants with no or low information need, (2) participants with moderate information need, and (3) participants with high information need. In

addition to the intensity of information need, a thematic analysis also revealed these different topics:

1. Need concerning communication with doctors
2. Need concerning information with thematic prevention/ precaution (including questions on nutrition and sports)
3. Need for information about medication
4. Need for health-related costs and their generation
5. Need concerning information about vital data and health parameters
6. Need for better exchange between health-related actors and institutions
7. Need for information about health insurance company
8. Need for age-related possibilities for obtaining information
9. Isolated further requirements for information that could not be classified.

In the following sections, the information needs for each of the 3 groups formed, broken down by topic, are presented.

Group 1: No or Low HIN

This group includes ID03, ID04, ID08, ID14, and ID16. ID04 and ID16 indicated that they had no need for or did not comment on health-related information. ID14 supported the statements of ID15, but otherwise did not express its own HIN (ID15 expressed moderate HIN in topic 2). Cancer screening was an important issue for ID03. In addition, ID03 had no current HIN at the time of the interview because the participant was generally satisfied with the information transfer of his own doctors and presented it as honest, open, and to the point. The HIN of ID08 depended on their own health situation. As this was satisfactory at the time of the interview, the participant had no particular HIN. However, the question of good prevention measures is interesting for ID08. ID08 also spoke of an experience in which her doctor was unable to answer all health-related questions. She then added information from the internet to the information she received. Both ID03 and ID08 were satisfied with their health situation and showed a need for prevention/precaution (topic 2).

Group 2: Moderate HIN

This group includes ID05, ID06, ID10, ID15, and ID18. All interview partners, except for ID10, addressed (topic 1) a need for communication with doctors. Thus, ID05 stated that she is generally satisfied with the transfer of information between the doctor and herself. In this context, it was perceived as negative that doctors do not have or cannot take enough time for the treatment, thus leading to treatment insufficiency. ID05 commented on the medication prescription as follows:

> *Before that, the doctors wrote down what the heart desires. They didn't bother at all, I think.*

However, ID05 trusted the doctors. ID06 was very satisfied with the medical expertise, the related organization, and accessibility of the information she encountered. This covered the largest part of ID06's HIN. Doctor appointments were dutifully documented by ID06:

> *This is very important for me. I always write down in my notebook how often I go to the doctor.*

It was important for ID06 that the information comes from a doctor and that the information is bundled with this doctor. ID15 was generally satisfied with the information he gets on health-related questions:

> *Well, information in general has never been withheld from me when I have asked for it [both at the doctor's office and on the internet].*

Access to information was given for ID15, even if it was sometimes problematic to obtain it because of a lack of exchange between doctors. In contrast, like ID05, ID18 considered the doctors' lack of time to be a problem:

> *The doctors only have three minutes' time. How is he supposed to explain [the meaning of the diagnosis: Morbus Sudeck] to me in three minutes.*

Due to the lack of time, information could not be transmitted sufficiently, which led to a compensatory measure in the form of information generation via the internet. Generally, ID18 was satisfied with the information about her health available to her. ID18 often relied on her own perception. In comparison with its sensation, ID18 ranked the quantity of information to be secondary.

ID05, ID06, ID18, and ID10, in particular, expressed an HIN on the topic of prevention/precaution (including questions on nutrition and sports; topic 2). For ID06, precaution was a very important issue, and ID18 stated that there is a need for further information. ID05, however, would have liked to get information on diet advice. ID10's HIN focused on sports activities. For example, ID10 needed instructions for his regular sports training sequences, which he has received from video recordings on a DVD. He had a need for daily alternating sports exercises and would have liked to know how far he can increase and vary his training. In this context, ID10 had an HIN to determine his sporting progress, too:

> *How has body fat percentage developed over time, for example?*

For ID05, ID06, and ID18, there was an HIN (topic 3) about the medication. ID06, in particular, focused on the interaction of medication. For ID06, the main trust in their own doctors was reflected in the way they deal with medication:

> *You don't even want to know what can be there. We rely on the doctors.*

Nevertheless, ID06 would have liked further information on medication and the reliability of medication effects, which is contradictory.

Furthermore, ID06, ID15, and ID18 had a (topic 4) need for health-related costs and how costs arise. ID15 would have liked to have more information about how hospitals and doctors bill patients. For ID18, the bills for clinical examinations and treatments were not transparent and comprehensible, which was why ID18 would have liked more information on this topic:

> *I do not understand the billing process, it is not comprehensible at all how they do it.*

ID06 was dissatisfied with the information received regarding health insurance coverage.

ID15 expressed the need for a better exchange between health-related actors and institutions (topic 6). Thus, ID15 was dissatisfied with the information flow between doctors and believed that the views of the doctors were not sufficiently congruent. For him, this was reflected in the diagnoses that were made. The fact that doctors make different diagnoses based on the same facts ensures that the information becomes more unreliable. ID15 would have liked to see more clarity, accuracy, and congruency from doctors. Besides, he was dissatisfied with the limited and inaccurate information flow from doctors and hospitals to their patients. He mentioned the example of a planned operation that was to be performed on him:

> *It was already three o'clock in the afternoon, when I was supposed to be picked up and I was still lying in the bed with my hospital gown open in the back. I thought I'd get going now. And I had not received the information.*

According to his own statement, after receiving no information, he checked out himself.

Moreover, ID15 would have liked more information on (topic 7) health insurance companies. Overall, ID15 was satisfied with the information he received from his health insurance company on the scope of services provided there. Accordingly, he would not consider it to be a problem that health information is being stored on his health insurance card if it was accessible to doctors, thus facilitating the exchange of information. ID15 found it interesting to know which information is stored in the health insurance card and which personal information can be viewed there.

ID18, however, had a need for age-appropriate means of obtaining information (topic 8). This was reflected in the desire for better guidance on the internet to obtain the desired information more quickly. Thematically interesting for ID18 were, among other things, hints for self-help groups to get reports of their experiences.

ID05 concluded by commenting on isolated further needs for information (topic 9). An HIN was defined here in terms of legal procedures, for example, toward companies. This was reported from a personal experience in which ID05 and ID04 became victims of a fraud during a coffee trip:

> *Yes, we were once badly fooled. We were on a coffee trip there and they sold us a product. [...] There were also people who said "yes, we did that too" and afterwards we found out that they mix people among the coffee trip participants who belong to them. Afterwards, you're always smarter.*

Group 3: High HIN

The group with high HIN encompassed ID02, ID07, ID09, ID13, and ID17. ID02, ID07, and ID17 had (topic 1) a need for communication with doctors. ID02 had the desire for more transparency in medical examinations. He was bothered by the fact that information only came when it was requested. In contrast, ID17 had a basic need for information regarding his or her health situation:

> *And I am someone, I said from the beginning, who wants to know what I have. I want to know how I have to handle it.*

For her, this handling of information was part of their information behavior. Furthermore, the personal relationship with the doctor was important for ID17, who was also the most important information source for her.

ID07, however, reported a recent experience that has had a lasting influence:

> *[My daughter] who is 27 weeks pregnant, will have twins, and her gynecologist said that she has to go to her family doctor. The family doctor said that the leukocytes were too high. And he, who then sent her home, said, "I'll check it out" and talked to her on the machine this morning and told her "yes, her gynecologist would get in touch with her next week."*

ID07 believed that doctors often lack the feeling for the context or the empathy for the patients' situation. She firmly believes that patients, especially her own daughter, have to put up with long waiting times and are informed relatively late about their own symptoms.

ID07 described the idea of the doctors' lack of empathy with the fact that doctors often have no intuition for someone not wanting further information. She talked about a procedure in which the flow of information led to her feeling nervous and restless:

> *That already strained me with what they said. "We'll saw your bone through there," and so on.*

She also described a third incident that had a lasting impact on her information needs. ID07 was much younger, and although she was still breastfeeding, this could have been dangerous for the child, as her treating physician had prescribed cortisone (Cortisone is a pregnane [21-carbon] steroid hormone. It is one of the main hormones released by the adrenal gland in response to stress).

ID07 considered this a wrong decision that originated from a lack of information generation:

> *Yes, and he sees that I have a child and does not ask me if I am breastfeeding or something like that, but prescribes me a cortisone medicine. You can't do that. That goes into the blood and then into the child.*

The described incidents led ID07 to the statement that she is not a doctor's friend and that she critically questions the information provided by them. Accordingly, she wanted several expert opinions on a diagnosis:

> *Somehow, I always have the feeling that I am missing information, because I say "yes okay, then I go to the next doctor. I'll ask his opinion about that. Or I'll ask a third doctor about this." That is, with one piece of information, I am therefore probably not so satisfied. Probably this will not be enough for me, then I would need a little bit more.*

Overall, the desire for credible doctors prevailed at ID07, as did the desire for self-determination:

> *You are sent from one doctor to another and don't have much of a choice to say: "But I'm going to see another doctor."*

ID02 and ID03 commented on (topic 2) the HIN on prevention/precaution (including questions on nutrition and sports). ID02 had an in-depth interest in cancer screening. ID13, in this regard, had an HIN on pain causes and management, as she was currently in pain. In connection with this, ID13 had an HIN that deals with muscle activity and performance. She was interested in how you can plan your daily activities meaningfully in that context.

ID13 and ID17 had (topic 3) an HIN about the medication. Thus, ID13 had an HIN on the effects and intake of medication. ID13 consolidated her own doctors to obtain information. ID17 showed a need for transparent communication in the field of medication. She found it difficult that she had to act as an information source when she visits a new doctor and had to inform him about the medication she was taking. ID02 and ID13 showed a need for (topic 4) health-related costs and their generation. For ID13, medical bills and the handling of costs by doctors and health insurance companies were mentioned as interesting points. ID02 was interested in questions that dealt with the composition of treatment costs. For example, ID02 stated that the flow of information between doctors and the patient on this subject was impersonal and inaccurate. For him, it was important to be able to understand the costs.

ID02, ID07, ID13, and ID17 commented on (topic 5) an HIN on vital data and health parameters. Thus, ID02 had the desire for direct clarification of available health data, for example, measured values. ID17 showed an HIN on health-related parameters and values and their personal significance. ID17 would have liked assistance in interpreting health-related data, as described in the first case of her pregnant daughter. ID13 indicated that vital signs were generally not very important to her. Blood pressure was excluded from this, even if ID13 stated that she was able to assess it well on the basis of body sensation:

> *I consider it very important, but I don't need to measure it, I can tell you by heart what it is like. [...]My blood pressure is always 140 over 80 with medication intake. [...] As soon as the lower value rises, I feel as if I really wanted to squabble.*

ID17 had a basic HIN; however, ID17 did not want to be reminded of her illness every day. This included, for example, the daily wearing of a measuring device, which she considered to be very stressful for her. Her wish for information and the desire not to be constantly reminded of her illness was somewhat of a dilemma.

ID02 and ID07 had a (topic 6) need for better exchange between health-related actors and institutions. For ID07, the flow of information between doctors and patients in this field was impersonal and inaccurate. She assumed that these processes were carried out by a third party, for example, a secretary, and that the attending physician was not even informed about the costs involved. In ID07's view, this matter was also an incomplete communication between doctors, which annoyed her personally.

On (topic 7) an HIN about health insurance companies, ID07 stated that it is a difficult matter on which she would have liked to have more information. In addition, ID13 stated that, according to her, health insurance companies work against, not with and for, patients, which leads to the exclusion of patients:

> *Something could come from the insurance company to make life easier for you... So that they'll be more active in approaching people.*

ID13 also expressed her opinion on (topic 8) the need for age-related means of finding information and would have liked to have easier access to it. The background was that, according to her statement, many older people do not have internet access.

Finally, ID07 and ID17 (topic 9) indicated isolated needs for further information. In ID07, it represented a desire for self-determination:

> *You are sent from one doctor to another and don't have much of a choice to say: I'm going to see another doctor.*

This led to an HIN about the availability of alternative doctors and a need for general information about doctors before a doctor becomes a patient. ID17 focused on the social environment and the HIN for relief measures. The reason for this need was her own heart disease manifested in the form of several heart attacks. According to patient ID09, there was an HIN for follow-up and preparation in addition to the discussion with the doctor. The doctor was considered to be the most important source of information for personal health information; however, ID09 felt that this is missing because of time pressure of the physician, lack of interest in the patient by the physician, or the fact that the physician does not take patients seriously:

> *They don't take you for full [...]I didn't understand at all what he said to me [...]I've written down the words (technical terms/unintelligible words), and I'm going to the family so a family member can translate them for me.*

For an appropriate exchange of information about health information, ID09 required the fulfillment of emotional and interpersonal needs by the physician as a prerequisite for exchanging information in a personal conversation. If a doctor did not comply, the need for information was covered by another source of information. Most importantly, she saw a difference between the specialists and the family doctor.

HIN Questionnaire Data

The descriptive results of the questionnaire data revealed that the information needs of older adults were quite low: 46% (12/26) of the sample were satisfied or rather satisfied 12% (3/26), 8% (2/26) were rather dissatisfied with the available health information, and 35% (9/26) were partly dissatisfied. HIN is indicated by the satisfaction with the information at hand. Information need was thus measured by how applicable participants considered the statement "*I am satisfied with the information I have available on health/my health*."

Information-Seeking Behavior Interview Data

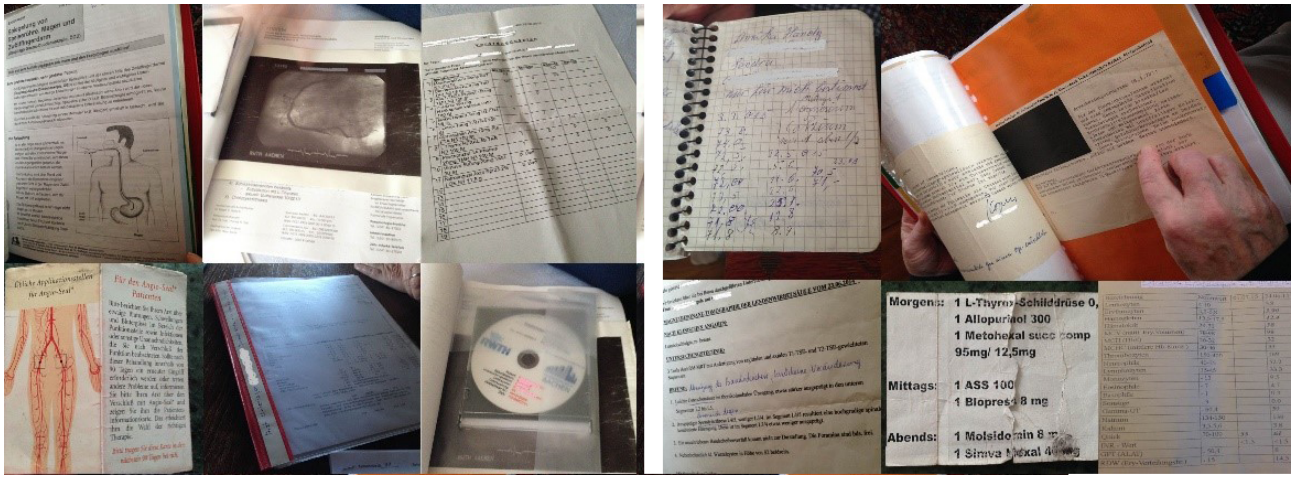
A total of 9 participants showed diverse tools that they already used to record, keep, and exchange information regarding their

health. Accordingly, Figure 1 documents pictures and screenshots. These artifacts included descriptions of surgical procedures; examination results of imaging procedures; tables with results of laboratory tests; folders with personal disease histories (prescriptions, diagnoses, findings, etc); handwritten medication overviews and appointment reminders; and personal diaries with medical data such as blood pressure, pain perception, and behavior.

Some participants clearly showed either an active (ID08, ID10, ID17, and ID18) or a passive (ID08, ID10, ID17, and ID18) information behavior. In contrast, ID06, ID07, ID13, and ID15 represented a mixture of active and passive information behaviors and were classified according to the statement into 1 of the 2 groups. ID03, ID04, and ID16 did not state anything about their information behavior. Participants with active information behavior most frequently conducted research on health issues (prevention) and searched for information about diagnoses and (risk of) examinations. Another important point describing active search was to actively exchange health-related information with the social environment and actively asking doctors (in the form of calls or personal conversations) if uncertainties or questions prevailed.

Participants with passive information behavior required a reminder to go to medical checkups instead of actively remembering or investigating information about it. Their passive information search behavior was reflected in the intake of medication. Here, the doctor played an active role in providing information on the effect and intake and giving further advice on the medication. Passive participants perceived information from the social environment instead of actively using information systems such as the internet to find information. The passive information retrieval process started by observing one's own symptoms before consulting a doctor. Some participants with passive information search behavior perceived health information as less desirable and irritating for some patients.

Figure 1. Participant's documentation of health-related information on paper and digital media.



Information-Seeking Behavior Questionnaire Data

Answers to the questions on information-seeking behavior showed that the largest part of the sample (7/26, 27%) used health information to change their health behavior. A total of 23% (6/26) used it to complement professional information either as preparation for a conversation or in addition to it (Table 3).

The results suggest that older adults use 2 to 3 sources to find relevant health information. On the basis of all valid answers, 15% (4/26) use 1 information source, 31% (8/26) use 2 sources, 35% (9/26) use 3 sources, 12% (3/26) use 4 sources, and 8% (2/26) use 5 different sources. All participants find health information on television shows. Here, 85% (22/26) of participants mentioned the German television show Visite (reports on medical topics). Other sources include newspapers and magazines (eg, *Apothekenrundschau*, a magazine distributed free of charge in German pharmacies) and the radio. On the basis of all valid answers, 73% (19/26) indicated that their information sources deliver the information they need, whereas 12% (3/26) said that using their information sources does not lead to the information they want. The other 12% (3/26) indicated that their information sources partly deliver the information they need. The majority of participants (20/26, 77%) were willing to share health-related information with other people. Only 19% (5/26) were unwilling to do so. The following table depicts more about how participants characterized their information acquisition (Table 4).

Table 3. Percentages of the different purposes older adults use health information for (N=26).

Use	Participants, n (%)
Generate knowledge	4 (16)
Change health-related behavior	7 (27)
Complement doctors' information	6 (22)
Making health decisions	2 (6)
Treatment	4 (17)
Exchange experiences	3 (11)

Table 4. Information-seeking behavior of the sample (N=26).

CISS[a]-item	Valid answers, n (%)					
	Applicable	Rather applicable	Partially applicable	Rather not applicable	Not applicable	Missing
"I am actively seeking information about (my) health."	6 (33)	2 (11)	4 (22)	0	6 (33)	8 (30)
"I am looking for information about health or my health rather casually."	5 (29)	1 (6)	0(0)	1 (6)	11(61)	8 (30)
"I am consciously looking for information about (my) health."	8 (44)	1 (6)	1 (6)	2 (11)	6 (33)	8 (30)
"I am passively looking for information about (my) health."	7 (39)	0(0)	3 (17)	0(0)	8 (44)	8 (30)
"I am permanently looking for information about (my) health."	2 (39)	0(0)	0(0)	1 (6)	15 (83)	8 (30)
"I am a curious person."	16 (64)	4 (16)	3 (12)	0(0)	2 (8)	1 (4)
"I am willing to take risks."	7 (28)	3 (12)	4 (16)	3 (12)	8 (32)	1 (4)

[a]CIS: Coping Inventory of Stressful Situations.

The Influence of Coping Strategies on HIN and HISB Interview Data

Task-Oriented Coping Strategy

The information needs of the participants (n=5) from the group with task-oriented coping behavior (T group) are clear and varied. For example, they include the exchange of information between health actors, transparent information on doctors and health insurance billing, and even nonexistent information needs. Thus, one assumption is that T group members are more interested in additional process-relevant organizational information. Information about one's own health is only needed and exchanged when complaints or symptoms occur. A clear diagnosis, cause, and precaution (eg, cancer precaution) and an exchange of experiences are desired. In addition, participant ID08 stated that information needs are primarily "*dependent on the health situation, and therefore are currently low*."

Across the entire group, doctors are viewed as the most important source of information, even if *the experiences were not always good*. The information behavior concerning one's own health is symptom-/illness-related and focused on the doctor. The internet is often mentioned but is critically viewed as a means of obtaining health information. For ID02, the lack of knowledge about a technology constitutes a hurdle for technology use, so that "*as far as health is concerned, I don't go there (comment: the internet) because I don't know how it works and I can't use it*." ID18's statement that "*A technical system doesn't replace the doctor. I mean, it can inform me, but the internet can't treat me*" reflects again the strong reliance on the doctor and shows that information is tied to treatment and action in general.

Avoidance-Oriented Coping Strategy

Only 1 female interviewee was assigned to the group with a pure avoidance-oriented coping strategy. The need for information of the interview partners with an avoidance-oriented coping strategy is comparatively low and relates to drug intake and effect. The doctor stands at the center of information retrieval. He/she has the patients' complete confidence, and information provided by him/her is not questioned or controlled. Great uncertainty and fear of all other sources exists because of fear of fraud or being taken advantage of, which results from personal experience. The doctor initiates any kind of information behavior. There is no individual drive to gain information. General and personal health information is obtained from a small number of information sources.

Task- and Avoidance-Oriented Coping Strategy

Only 1 female interviewee was assigned to this group. Her need for information is similar to the needs of group A: she seeks exclusive information about occurring diseases or complaints that are completely provided by doctors. Similarly, the search for information behavior is only active in the case of complaints and then directed solely at doctors. Information available on television is randomly included in the current personal situation. There is high distrust of all sources of information not related to doctors.

Task- and Emotional-Oriented Coping Strategy

Each participant in the TE group demonstrated a high need for information. Compared with the participant in group A, this includes just as much diversity, but in the TE group, it focuses much more on the individual than on the indirectly related organizational processes. Moreover, compared with group A, interest is not linked to a disease or symptom but is generally present. To obtain health-related information and data, fitness trackers, blood pressure monitors, and digital training instructions are used to document and independently influence one's own health. Doctors are mentioned as the most important source, but "*the different diagnoses which one receives from doctors to one and the same symptom show...that one should remain critical toward doctors*" (ID14). In no other group is supplementing medical information with active, personal information gathering so self-evident: "*I always in-form myself in advance before I go to the doctor*." Additionally, it is no wonder that, compared with other groups, emotional states play a major role here: "*A personal relationship (to the doctor) is very important*" stated ID17 and described trust to be an important factor by stating that "*I find it pleasant when you can see the person directly, look him in the eye. That creates trust*."

Task-, Emotion-, and Avoidance-Oriented Coping Strategy

Of the 3 interview participants in the TEA group, only 1 woman gave detailed answers to the interview questions. As with the participant in group A, doctors are viewed as responsible for providing information about the participant's health. However, ID07 stated:

> *those (doctors) unfortunately often lack the feeling for the context. They lack empathy for the situation of uncertainty in which the patient finds himself.*

Here, the method of information transfer is primarily criticized, which does not take sufficient stock of the patient's emotional world: "*You feel dispatched and inadequately treated*" (ID07). Unlike the TE group and similar to the T and A groups, the required information includes diagnoses and medication information and help to interpret laboratory findings and treatment methods. These are obtained without exception from personal sources of information such as the doctor or pharmacist or, in exceptional cases, from a medically trained relative. The remaining male interviewees of the TEA group indicated that they did not want to know about health or their personal health.

The mixed methods analysis of normalized code frequencies in the separate CISS groups supports the preceding qualitative view. The results should be viewed against the background of the group size (A: n=2, T: n=7, TE: n=9, TA: n=2, and TEA: n=4). Codes concerning information behavior most frequently occurred in the TE group.

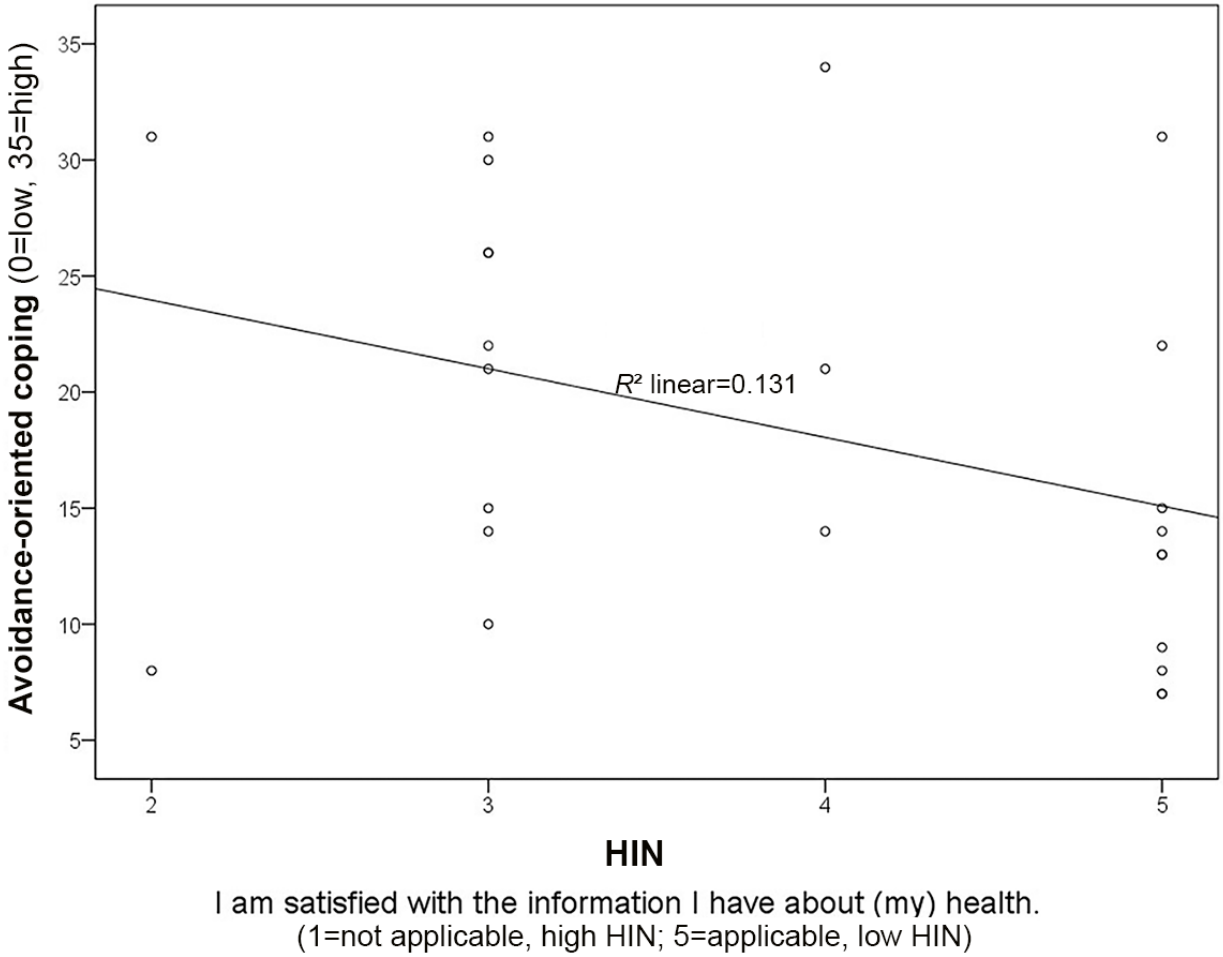
In short, it can be stated that qualitative interviews suggest an HIN influenced by individual coping strategies. This matches the results of international researchers and theoretical models [15,17,19,21]. Particularly noticeable in the quantitative analysis of the qualitative data was the influence of the avoidance-oriented coping strategy (group A). People who applied the avoidance strategy entirely or partly had a descriptively much lower HIN. This seems to be intrinsically motivated because satisfaction with the doctor was not necessarily accompanied by an increased HIN. The TE group was the most open to technology use and the collection and interaction of and with its own health-related data. Further investigations could serve to identify factors that explain this observation beyond coping strategies.

The Influence of Coping Strategies on HIN Questionnaire Data

The score of participants' satisfaction with information at hand (ie, information need; $D(26)=0.294$; $P<.001$) is significantly different from normality. The numerical scores of each dimension were as follows: task-oriented coping $D(25)=0.158$, $P=.11$; emotion-oriented coping $D(25)=0.114$, $P=.25$; and avoidance-oriented coping $D(24)=0.193$, $P=.80$. The bootstrapping method and bivariate correlation models were applied to investigate the relationship between the scores of individual coping strategies and information need.

Bias-corrected and accelerated bootstrap 95% CIs (95%, BCa CI) are reported in square brackets. No relationship was found between information need and the task-oriented coping strategy score ($r_s=-0.056$, 95% BCa CI −0.469 to 0.445; $P=.79$). In addition, no correlation was found between information need and the emotion-oriented coping strategy score ($r_s=-0.149$, 95% BCa CI −0.532 to 0.260; $P=.49$). However, a trend was found for a negative correlation between the avoidance-oriented scale and information need ($r_s=-0.378$, 95% BCa CI −0.730 to 0.092; $P=.05$). The more strongly a person is characterized by an avoidance-oriented coping strategy, the lower is the person's health-related information need (Figure 2).

Figure 2. Relationship between the scales of HIN and avoidance-oriented coping. HIN: Health Information Need.



The Influence of Coping Strategies on HISB Questionnaire Data

Emotion-oriented coping has a negative relationship with the perceived success of an information source (r_s=−0.607, 95%, BCa CI −0.876 to −0.139; *P*=.20). The more people rely on the emotion-oriented coping strategy, the lower they rated the success of the information they received from the sources they used. The casualness with which a person looks for health-related information is positively related to the success of an information source (r_s=−0.620, 95% BCa CI 0.302 to 0.884; *P*=.01). The more casually a person searches, the higher the person rates the source in terms of success. The consciousness with which a person conducts the information search is positively correlated with the person's activity level in the search (r_s=0.929, 95% BCa CI 0.839 to 0.982; *P*<.001). The more consciously a person searches, the more active will be the search. The consciousness with which a person conducts the information search is positively correlated with the avoidance-oriented coping strategy (r_s=0.561, *t* 95% BCa CI 0.056 to 0.889; *P*=.03). The more consciously a person searches for health-related information, the higher the scale value of the avoidance-oriented coping strategy is.

Discussion

Principal Findings

This study on HINs and HISBs of older adults investigated the general context of data visualizations in a group of 18 adults aged between 50 and 91 years. Interviews on the topic of HINs and HISBs were conducted, transcribed, coded, and qualitatively analyzed. Questionnaires on social demographics and coping strategies served as a basis for comparing qualitative and quantitative results. Essentially, the results indicate a heterogeneous need for information on the part of older people, where one part of the population needs and desires the exchange of personal health data and the other part adopts an attitude of avoidance. There is a need to deal with one's own health data as a supplement to professional and medical sources of information.

Discussion of RQs

The first RQ to be answered in this regard was RQ 1: *Which HINs do older adults have?* The results indicate that the health-related information needs of the older people surveyed are not homogeneous. More than half of the participants were satisfied with the information available. According to the definition by Case et al [15], this corresponds to a low need for information. The majority of the interviewees justified the need for health information by stating that they had no health complaints. Another reason provided by the participants in this

group was the unsubstantiated assumption that the acquisition and examination of information about one's own health can lead to the triggering of diseases or an increase in the current pain. This statement is contradictory because a need for preventive, relevant information was often formulated at the same time. This contradiction can point to a need for a more detailed consideration of individual types of health-related information. This study examined the general need for information and revealed that different types of health-related information appear to have different effects on patients and their behavior. In addition to the group that has little or no need for information, there is also a group who makes intensive efforts to gain information about their health. For these people, it is not enough to know what is necessary; rather, they use additional sources of information such as the internet or television to supplement the information they receive from their doctors. For these individuals, information acquisition is considered part of their coping strategy. Only if these patients are sufficiently informed about their illness, do they consider themselves able to make decisions and communicate with doctors.

It was particularly surprising that half of all respondents already collected health data on paper and were using computers. These included notes listing the type and quantity of medication patients carry in their wallets to provide a basis for decision making. Furthermore, doctors had made laboratory results and examination values available to the patients in tabular printouts. Older people had disease histories meticulously collected in folders consisting of examination results, x-rays, medication instructions, and accounts together with detailed visualizations of surgical procedures. Occasionally, participants documented blood pressure, sports activities, and symptoms in a digital form or wrote pain diaries to draw conclusions about causes and adapt their health-related behavior accordingly. The interviewees were among the generation that did not grow up with digital technology. It can, therefore, be assumed that the described number of people digitally documenting their health will increase even further with the growing number of *digital natives*. Here, it is necessary to investigate whether and to what extent the group of information avoiders will play a role in digital health systems that visualize personal health data.

Results regarding health information relevant to older people indicate that more information is needed concerning preventive measures and everyday healthy behavior. Most importantly, there was a lack of a decision-making basis for one's own behavior. This conclusion is supported by the information regarding the objectives pursued with the collected information. About one-third of the participants stated that they wanted to change their own health behavior according to the information they had collected. The need for information coming directly from the doctor and more intensive communication between doctors and patients reveals the fundamental importance of doctors for older patients. Despite the doctors' position as the most trustworthy source of information, older people see their lack of time as a barrier to having their information needs met. Most importantly, current billing structures do not allow, for example, detailed clarification of medical terms from the treating physician or receipt of more treatment and diagnosis-specific information from the doctor. Even if digital health systems have the potential to compensate for the doctors' aforementioned lack of time, the doctors' acceptance of digital health systems would require a clear billing concept for services provided digitally.

With regard to RQ 2, *How do older adults acquire needed health information?*, the results indicate that health issues and symptoms initiate the information search. Furthermore, if the principal information source—the doctor—does not provide enough information, search activities are initiated. This is in line with the model of information-seeking behavior by Wilson, which states that the failure of one source to provide information motivates search activities. Older adults' information behavior can thus be considered occasional. Occasional searching could be an alternative explanation for the heterogeneous need for general information. The health status of participants was not explicitly investigated and needs to be considered in future studies on this topic. Furthermore, when it comes to health, the results suggest that the most frequently used and most trustworthy information source is the doctor. These results are consistent with those of age-independent studies. In particular, the older adults attribute medical competence only to the doctor; therefore, they put the decision about a treatment completely in the hands of the doctor. At the same time, similar to the results of the study by Geuter and Weber [7], trust is perceived as a particularly important determinant that arises from personal contact with the treating physician. In addition to the professional competence of the source, the influence of emotional factors on the search for information becomes evident. Besides doctors, television and even the internet are sources of information.

With regard to RQ 3, *How does the coping of older adults relate to their HINs/HISB?*, it can be stated that the qualitative and quantitative results indicate an influence of coping strategy on HIN and HISB. An avoidance-oriented coping behavior especially leads to a lower information need. Assuming that a lower information need results in avoiding information search with technology, avoidance-oriented coping behavior can be considered as a hurdle for health technology and health data visualization use. Further investigations on system design with regard to coping strategies are needed.

Limitations

The main limitation of this study is its ecological validity. Its cross-sectional design provides insights into HISBs of older adults for a single point in time. Although it closes the research gap of investigating how the specific population of older adults requires information, questions on HINs with respect to ongoing health conditions remain unanswered. As with many studies on the need for health and patient information, the subjective character of this study might be subject to social desirability. Older adults, in particular, often feel the need to conclude from the questions what might be expected from them to adjust their answers accordingly. Future studies on HIN would benefit from controlling this variable. Ecological validity could be improved by investigations within a natural setting, meaning that data from patients' providers or search engines could be analyzed to triangulate subjective data using objective observational data, interaction, and logfiles.

Furthermore, the study was conducted with participants from Aachen, Germany. Germany's socioeconomic parameters such as health care, average monthly income, life expectancy, education level, and density of dental care provision match those of other European Union (EU) member countries. Consequently, the data collected should be comparable with those of most of the other EU members, but HINs differ in countries with different economic or cultural backgrounds. At the same time, the results should be subjected to a generalization against the background of sample size and the procedure for acquiring participants.

Another factor that may have influenced the results of the study is that the patients interviewed may have a different understanding of terms in relation to the queried criteria than initially assumed. Against the background of this study, it seems quite probable that the criteria asked for, such as the *health-related information need*, for example, could be understood differently than initially assumed. The patients only answered and assessed these questions on a personal level. In addition, the study participants felt that the question regarding their satisfaction with the information available to them was in part an evaluation of their physician because they understood their physician to be responsible for communicating this information. Their relationship and experience with their physician is thus an influencing factor.

Conclusions

The results regarding the general need for information identify the need for older people to gain insight into personal health data and to use this as a basis and addition to medical information provided by physicians. This motivates successive investigations on age-differentiated, ergonomic considerations of the visualization of personal health data. However, it should be noted that not every older adult wishes to independently analyze his/her own health data. When provided with health-related data, participants most importantly require support to interpret the data and assess their significance for their personal situation. As the daily use of health-related data puts the disease first, a visualization of data with a direct reference to the disease carries the risk of reduced acceptance, adherence, and use of the corresponding system. In contrast, the need for behavior-changing and preventive measures suggests that data visualizations that allow for conclusions about personal behavior and its correlation with symptoms might positively influence these factors. One unresolved issue in this regard is the extent to which data visualization can increase the motivation of the patient to change a behavior.

This study focused on the need for health information to examine the broad context of digital health systems. Empirical evidence for a correlation between health-related information needs and data visualization/use of technology is lacking and needs to be investigated, especially for the group of older adults. There is also little empirical knowledge about the importance of trust in connection with the visualization of personal health data. Although investigations on factors that influence or generate trust in the data might also be especially relevant for the health care domain, it is still unclear how trust develops in the context of health-related decision making and how corresponding processes proceed or if data visualizations might even increase the user's acceptance of a digital health system.

Finally, it remains to be clarified whether there is a difference between the information needs of chronically ill and acutely ill people or whether differences in HIN arise predominantly according to observed symptoms or life and care experience. Particularly vulnerable groups (those with Parkinson disease, Alzheimer disease, etc) must be taken into account.

Implications for Technology Development

The following implications for technology developments were derived from presented results:

1. Digital health technology might be more accepted if its use is recommended and accompanied by the physician.
2. The occurrence of personal symptoms and diagnoses might trigger individual information search behavior.
3. The physician should provide information on illness and medication more effectively, whereas the patient provides information on health-related parameters to the diagnostic process most effectively.
4. Digital systems that can support the patient in everyday documentation of symptoms and complaints to support the diagnostic processes of doctors are required.
5. Patients need support in documenting symptoms and complaints.
6. Cooperation and data exchange of all actors involved in diagnosis and treatment simplifies this for the patient.
7. Comprehensibility and competence of the information source is a key requirement of the patient and should therefore be considered in system development.
8. Adaptive systems for coping strategies are required to address the nonhomogeneous health-related information needs of older adults; therefore, digital health systems must enable patients and users to assess the trustworthiness of information and develop trust.
9. Older adults require diverse types of health-related information and use different methods to acquire information. The planning and development of digital health systems should combine and harmonize different sources of information. Not only user groups with their skills and abilities but also the characteristics of the information sources should be taken into account to effectively coordinate their interaction.
10. Communication strategies implemented in the system that put health rather than illness in the foreground foster acceptance and adherence.

Regarding the design recommendations, it must be considered that these will have to be validated before being actually applicable to system design.

Acknowledgments

This publication is part of the *TECH4AGE* research project funded by the Federal Ministry of Education and Research, under grant no. 16SV7111) and promoted by Verein Deutscher Ingenieure Innovation+Technik GmbH.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaires.
[PDF File (Adobe PDF File), 183 KB - humanfactors_v8i1e15858_app1.pdf]

References

1. Marć M, Bartosiewicz A, Burzyńska J, Chmiel Z, Januszewicz P. A nursing shortage - a prospect of global and local policies. Int Nurs Rev 2019 Mar;66(1):9-16. [doi: 10.1111/inr.12473] [Medline: 30039849]
2. Tresp V, Marc Overhage J, Bundschus M, Rabizadeh S, Fasching PA, Yu S. Going digital: a survey on digitalization and large-scale data analytics in healthcare. Proc IEEE 2016 Nov;104(11):2180-2206. [doi: 10.1109/JPROC.2016.2615052]
3. Lintern G, Motavalli A. Healthcare information systems: the cognitive challenge. BMC Med Inform Decis Mak 2018 Jan 11;18(1):3 [FREE Full text] [doi: 10.1186/s12911-018-0584-z] [Medline: 29325563]
4. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1) [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
5. Soergel D. Annual review of information science and technology. Library Inform Sci Res 2003 Mar;25(1):111-113. [doi: 10.1016/s0740-8188(02)00179-2]
6. Jenkins V, Fallowfield L, Saul J. Information needs of patients with cancer: results from a large study in UK cancer centres. British Journal of Cancer Research 2001 Jan 5;84(1):48-51. [doi: 10.1054/bjoc.2000.1573] [Medline: 11139312]
7. Geuter G, Weber J. Informationsbedarf Chronisch Kranker Menschen Bei Der Krankenhauswahl: Untersucht Unter Besonderer Berücksichtigung Des Internets. Institut für Pflegewissenschaften. 2009 Jan. URL: https://www.universitaet-bielefeld.de/gesundhw/ag6/downloads/ipw-140.pdf [accessed 2020-12-28]
8. Sohn T, Li K, Griswold W, Hollan J. A Diary Study of Mobile Information Needs. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2008 Presented at: CHI'08; May 2-4, 2008; Florence, Italy p. 433. [doi: 10.1145/1357054.1357125]
9. Tate T, Russell-Rose T. The Information Needs of Mobile Searchers: A Framework. In: Proceeding of Searching4Fun Workshop. 2012 Presented at: ECIT'12; April 1, 2012; Barcelona, Spain. [doi: 10.1016/b978-0-12-396981-1.00008-2]
10. Lambert SD, Loiselle CG. Health information seeking behavior. Qual Health Res 2007 Oct;17(8):1006-1019. [doi: 10.1177/1049732307305199] [Medline: 17928475]
11. Wilson T. Human information behavior. Inform Sci J 2000;3:49-56. [doi: 10.28945/576]
12. Wilson TD. Activity theory and information seeking. Annual Rev Inform Sci Technol 2008;42(1):119. [doi: 10.1002/aris.144.v42:1]
13. Theis S, Schäfer D, Bröhl C, Schäfer K, Rasche P, Wille M, et al. Predicting technology usage by health information need of older adults: Implications for eHealth technology. Work 2019;62(3):443-457. [doi: 10.3233/WOR-192878] [Medline: 30909259]
14. Miller SM, Mangan CE. Interacting effects of information and coping style in adapting to gynecologic stress: should the doctor tell all? J Pers Soc Psychol 1983 Jul;45(1):223-236. [doi: 10.1037//0022-3514.45.1.223] [Medline: 6886967]
15. Case DO, Andrews JE, Johnson JD, Allard SL. Avoiding versus seeking: the relationship of information seeking to avoidance, blunting, coping, dissonance, and related concepts. J Med Libr Assoc 2005 Jul;93(3):353-362 [FREE Full text] [Medline: 16059425]
16. Yagil D, Cohen M, Beer JD. Older adults' coping with the stress involved in the use of everyday technologies. J Appl Gerontol 2016 Feb;35(2):131-149. [doi: 10.1177/0733464813515089] [Medline: 24652919]
17. Wilson T, Walsh C. Information Behaviour: an Inter-disciplinary Perspective: a Review of the Literature. British Library Research and in-novation Centre London. British Library. 1996. URL: http://www.worldcat.org/title/information-behaviour-an-inter-disciplinary-perspective-a-review-of-the-literature/oclc/36084563?page=citation [accessed 2021-01-01]
18. van Zuuren FJ, Wolfs HM. Styles of information seeking under threat: personal and situational aspects of monitoring and blunting. Pers Individual Diff 1991 Jan;12(2):141-149. [doi: 10.1016/0191-8869(91)90097-u]
19. Kim K, Lustria ML, Burke D, Kwon N. Predictors of cancer information overload: findings from a national survey. Inf Res 2007;12(4):-.

20. Kuckartz U. Realizing Mixed-Methods Approaches with MAXQDA. Philipps-Universität Marburg. 2010. URL: https://www.maxqda.com/download/MixMethMAXQDA-Nov01-2010.pdf [accessed 2001-12-28]
21. Warner E, Murray A, Palmour V. Information Needs of Urban Residents. Final report of the Baltimore Regional Planning CouncilWestat Inc Washington, DC: US Department of Health, Education and Welfare. 1973. URL: https://eric.ed.gov/?id=ED088464 [accessed 2021-01-01]
22. Endler NS, Parker J. Coping Inventory for Stressful Situations (CISS). Multi-Health systems Incorporated. 1990. URL: https://www.pearsonclinical.co.uk/Psychology/ChildMentalHealth/ChildMentalHealth/CopingInventoryforStressfulSituations(CISS)/PDFReports/Technical.pdf [accessed 2021-01-01]
23. Endler NS, Parker JD. Assessment of multidimensional coping: task, emotion, and avoidance strategies. Psychol Assess 1994 Mar;6(1):50-60. [doi: 10.1037/1040-3590.6.1.50]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]

Abbreviations

CISS: Coping Inventory of Stressful Situations
EU: European Union
HIN: health information need
HISB: health information–seeking behavior
RQ: research question
TA: task-avoidance oriented coping
TE: task- emotion oriented coping
TEA: task-emotion-avoidance oriented coping

Edited by B Price; submitted 13.08.19; peer-reviewed by R Kelly, R Islam; comments to author 03.10.19; revised version received 27.02.20; accepted 04.04.20; published 01.02.21.

Please cite as:
Theis S, Schäfer K, Schäfer D, Jochems N, Nitsch V, Mertens A
The Relationship Between Individual Coping and the Need to Have and Seek Health Information Among Older Adults: Exploratory Mixed Methods Study
JMIR Hum Factors 2021;8(1):e15858
URL: https://humanfactors.jmir.org/2021/1/e15858
doi:10.2196/15858
PMID:33522973

Original Paper

Effects of User Characteristics on the Usability of a Home-Connected Medical Device (Smart Angel) for Ambulatory Monitoring: Usability Study

Noémie Chaniaud[1], PhD; Olga Megalakaki[1], PhD, Prof Dr; Sophie Capo[1], PhD; Emilie Loup-Escande[1], PhD

Centre de Recherche en Psychologie: Cognition, Psychisme et Organisations, Université de Picardie Jules Verne, Amiens, France

Corresponding Author:
Noémie Chaniaud, PhD
Centre de Recherche en Psychologie: Cognition, Psychisme et Organisations
Université de Picardie Jules Verne
Chemin du Thil
Amiens, UR UPJV 7273
France
Phone: 33 3 22 82 70 59
Email: noemie.chaniaud@u-picardie.fr

Abstract

Background: The Smart Angel home medical device allows ambulatory surgery patients to monitor their own health by taking their blood pressure and oxygen levels and answering a health questionnaire from home. Currently, this device is a prototype in the design phase, and no usability evaluation has been performed. This preventive device must be usable by patients with different profiles; however, it is important to select patients carefully to ensure their safety when using the device. As such, it would be interesting to know how to select or exclude patients. However, the links between user characteristics and the usability of this home medical device remain unclear.

Objective: This study aims to better understand the links between certain characteristics of potential patients (ie, age, education, technophilia, and health literacy) and the usability (ie, effectiveness, efficiency, and satisfaction) of Smart Angel, as defined by the ISO 9241-11.

Methods: We conducted an experimental study involving 36 participants investigating the effects of 4 patient characteristics (ie, age, education, technophilia, and health literacy) on usability, measured in terms of effectiveness, efficiency, and satisfaction. A mixed methods approach (subjective vs objective) using a variety of standard instruments was adopted (direct observation, video analysis, and questionnaires). First, to help participants project themselves into the real use of the Smart Angel device, they watched a scenario in a video. Second, the participants completed a set of questionnaires to show the extent of their health literacy level (Newest Vital Sign [NVS] and the Health Literacy Survey [HLS]) and then operated Smart Angel devices. Efficiency (ie, handling time) and effectiveness (ie, number of handling errors) measures were collected by video analysis. Satisfaction measures were collected by a questionnaire (System Usability Scale [SUS]). The qualitative observational data were coded using inductive analysis by 2 independent researchers specialized in cognitive psychology and cognitive ergonomics.

Results: The results show a moderate and positive correlation between age and effectiveness (r=0.359; *P*=.03) and efficiency (r=0.357; *P*=.03). There is strong correlation between health literacy scored by the NVS and effectiveness (r=0.417; *P*=.01), efficiency (r=-0.38; *P*=.02), and satisfaction (r=0.45; *P*=.006). However, there is a weak correlation between technophilia and usability and no relationship between education level and usability.

Conclusions: Our results show that literacy level and age are 2 important factors to consider when selecting future users of the Smart Angel device to ensure patient safety. This study also serves as an example promoting mixed methodologies in assessments of medical device usability that cannot be performed under real-world conditions.

(JMIR Hum Factors 2021;8(1):e24846) doi:10.2196/24846

KEYWORDS

user characteristics; health literacy; home medical devices; usability study; remote health; home health

Introduction

Background

Outpatient surgery has been on the rise in recent years. Performed operations are increasingly complex and dangerous for patients who have to manage their convalescence at home. The Smart Angel device is a home-connected medical device specifically designed to prevent postsurgical complications related to outpatient surgery. The purpose of this device is to facilitate the patient's return home by maintaining a link with the hospital. Upon returning home after an operation, the patient is required to use the device to send all their vitals 3 times a day for 1 week before returning the equipment to the hospital center. This postoperative follow-up may also enable patients to manage their convalescence better by avoiding all-too-frequent returns to emergency services or outpatient consultations [1].

Currently, this system is in an early design stage. Like any medical device, this tool must follow safety and quality standards [2] and usability standards [3] to meet the requirements of European Conformity (CE marking) for marketing. However, even today, the deployment of these connected medical devices is still hindered by their complexity of use, directly implying a lack of usability [4-6], thus impacting patient safety. With this in mind, Kortum and Peres commented, “A lack of usability may cost lives” [7].

Usability

Usability is defined by the ISO 9241-11 [3] as “the degree to which a product can be used, by identified users, to achieve defined goals in an effective, efficient, and satisfactory manner, within a specified context of use.” This concept, which is still discussed by the scientific community, has 3 distinct dimensions: (1) *effectiveness:* the accuracy and completeness with which users achieve certain objectives; (2) *efficiency*: the relationship between accuracy and the resources used to attain it; and (3) *satisfaction*: user comfort and a positive evaluation of user interaction. Defined by these 3 dimensions, usability is linked to its context of use, characterized by 4 components: the task, the environment, the resources, and the users [3].

Despite the use of methodologies that involve the user in the design process [8-10], usability problems persist. There are 2 arguments in the literature that may explain this finding: (1) the lack of a standardized framework and method in usability studies [11-15], and (2) a lack of knowledge of the impact of the use context [16,17] on usability, in particular, user characteristics.

User Characteristics

Age, Level of Education, Technophilia, and Health Literacy

Several researchers have recently investigated the relationship between user characteristics and the usability of connected devices in health care [11,18-22]. In particular, 4 user characteristics have been studied in the scientific literature: age [11,20,22-26], level of education [11,19,20], technophilia (ie, experience in information technology and previous experience with medical devices [11,23,27]), and health literacy [20,24,28,29]. In most studies, authors tend to agree on these interrelationships when investigating different devices. We detail these studies below.

Age

Many authors have examined the influence of age on the usability of connected devices in health care. Most of these authors concur on the influence of age on usability. For example, Georgsson and Staggers [11] investigated the usability of a diabetes management app running on a smartphone using the metrics of ISO 9241-11 (effectiveness, efficiency, and satisfaction). The authors found that the younger age group (30-49 years old) made fewer errors (ie, was more *effective*), was faster (ie, more *efficient*), and more *satisfied* (System Usability Survey [SUS] score of 88.33 vs 77.14) than the older group (50-69 years old). Sparkes et al [23] examined the usability of remote cardiac testing and found that the age of the participants impacted their ability to install the equipment. Younger subjects appeared to be more comfortable than older subjects. Jones and Caird [25] examined the use of a blood glucose meter and found that younger subjects had fewer difficulties and made fewer errors (ie, were more effective) than older subjects. Mykityshyn et al [26] also examined the use of a glucometer and found that young subjects were faster (ie, more efficient) than older subjects, regardless of the instruction format provided (written and drawn vs video). Van der Vaart et al [20] evaluated the usability of an application for monitoring the symptoms of 32 narcoleptics and found that usability (measured in terms of the number of tasks completed and problems encountered) was moderately and positively correlated with age and eHealth literacy level.

However, Liang et al [19] found no relationship between age and satisfaction as measured by the SUS score in their study on the evaluation of 7 health devices used by the general public (eg, connected watches), conducted with a sample of 388 participants. Similarly, Jensen et al [18] found no relationship between usability and the age of participants with respect to access and use of online health information. The authors explain that this result is probably due to the contrast in health literacy levels that would have taken precedence over the other variables.

Level of Education

The level of education is also a variable found in many usability assessments. However, to our knowledge, no studies have proven this link. Georgsson and Staggers [11], Liang et al [19], and Van der Vaart et al [20] have all found a lack of association between participants' level of education and usability (ie, effectiveness, efficiency, and satisfaction).

Technophilia

Differing results have been reported regarding the influence of technophilia—experience with information technologies (IT) and previous experience of medical devices—on usability. Georgsson and Staggers [11] found that those with more technology experience (what the authors call “IT/computer experience”) made fewer errors (ie, were more effective), were faster (ie, more efficient), and were more satisfied with the diabetes management application (+5 points for the SUS score). Conversely, Harte et al [27] conducted regression analyses between technology experience and SUS score and found no

significant effect when evaluating a smartphone health app. Finally, Sparkes et al [23] showed that familiarity with the technologies seemed to have an influence on the correct installation of their device.

Health Literacy

Definition and Assessments

Health literacy is a user characteristic that can be expected to influence medical device usability [18,20,28,30]. Due to its multidimensionality, however, this characteristic is complex to define and difficult to assess. Sørensen et al [31] describe it as "an individual's knowledge, skills, motivation, and ability to identify, understand, evaluate, and use health information in decision-making in health care, disease prevention, and health promotion to maintain or improve lifelong quality of life." However, this notion is often mentioned as a determinant to be considered in therapeutic education [32], prevention [33], therapeutic adherence, access to health information [18], and even recovery rate [32,34]. However, to our knowledge, no study has assessed the level of health literacy among the French population at the national level.

In terms of evaluation, health literacy is particularly difficult to measure for at least two reasons. The first reason concerns its multidimensional specificity [31]. The second reason is that health literacy is not related to socioeconomic criteria as might be intuitively assumed [35].

Currently, there are 2 main methods of measuring health literacy [36]: (1) questionnaire methods, by which an individual's abilities are assessed, and (2) self-reported methods, by which an individual's behaviors towards a health professional are directly observed. Currently, few tools exist in the French language compared to the 51 English-language instruments identified by Haun et al [37]. The most frequently used and cited instruments are part of questionnaire-based methods; they are the Test of Functional Health Literacy in Adults (TOFHLA) [38], the Rapid Estimate of Adult Literacy in Medicine (REALM) [39], the European Health Literacy Survey (HLS-EU) [31], and the Newest Vital Sign (NVS) [40]. However, these instruments have several limitations. Among these instruments, the REALM is more like a reading test than a comprehension test since participants are asked to read medical terms. The short version of the TOFHLA (ie, the Short Test of Functional Health Literacy in Adults [S-TOFHLA]), which assesses respondents' level of comprehension, seems more adapted to Swiss culture than to French culture [41] (indeed, direct reference is made to the Swiss health insurance system and the transmission of certain documents that do not apply to the French social security model). In addition, the validity of S-TOFHLA is currently the subject of some controversy due to inconsistencies in the interpretation of its component items [42]. Another instrument proposed in the literature, the NVS [40], shows a strong correlation (Cronbach α>.76) with the measurement of S-TOFHLA [43]. It also assesses some of the respondents' cognitive skills (reading, writing, comprehension, numeracy). Finally, the HLS-EU is based on the multidimensional literacy model of Sørensen et al [31]. This tool has identified important gaps in 8 European countries, as approximately 1 in 2 people reportedly have a problematic or inadequate level of health literacy [44].

Health Literacy and Usability

In the context of health technologies such as connected medical devices, which are increasingly becoming part of patient life, studies on the correlation between health literacy and usability are still rare or exploratory. Monkman and Kushniruk [21] propose an assessment of usability by considering health literacy through the design and validation of heuristic criteria. To do so, the authors adapted a set of existing guidelines for designing health-specific websites to make the content more understandable to users with a reliable level of health literacy. Using an electronic personal health record system, Czaja et al [28] were able to show that populations with low literacy levels had more difficulty using these tools. Kim and Xie [29] conducted a systematic review of articles examining the impact of low health literacy on the use of eHealth devices. Based on 74 studies, the authors conclude that the major barrier to accessing and using online health information for individuals with low literacy is strongly related to website usability. Jensen et al [18] found that participants with low levels of health literacy (as measured by REALM) used health technologies less. Those with low levels of health numeracy (as measured by TOFHLA) would have limited access to these technologies. This latter finding is consistent with those of Kaufman et al [24], who also concluded that low numeracy could be a barrier to using a telemedicine system. Chaniaud et al [30] showed that it is necessary to obtain a minimum level of prior health knowledge to use home medical devices. Finally, to our knowledge, no experimental studies have empirically characterized links between health literacy and usability in terms of efficiency, effectiveness, and satisfaction.

Study Objective

We have seen that the complexity of using medical devices resides essentially in usability problems [29], all the more so as they must be usable by patients with diverse profiles. In this sense, consideration of user characteristics, including age, education, technophilia, and health literacy, are important factors to consider in the design of a connected medical device such as Smart Angel for a patient's home. However, to our knowledge, no study involving all 4 of these characteristics has been conducted. Moreover, the relation between these characteristics and usability remains unexplored in the literature. Thus, the aim of this paper is to better understand the relationships between the 4 user characteristics of age, educational level, technophilia, and health literacy, and the usability (measured by effectiveness, efficiency, and satisfaction) of a connected medical device intended for a patient's home.

To do this, we formulated 4 hypotheses: (H1) older users will be less *effective*, *efficient*, and *satisfied* with the Smart Angel connected medical device than younger users [11,20,25,26]; (H2) users with a low level of technophilia (IT and medical device experience) will be less *effective*, *efficient*, and *satisfied* with the Smart Angel connected medical device than those with a high level of technophilia [11,23,27]; (H3) the level of education will not affect the *effectiveness*, *efficiency*, and *satisfaction* with the Smart Angel connected medical device

[11,19,20]; and (H4) users with low levels of health literacy (as measured by NVS and HLS-EU scores) will be less *effective*, *efficient*, and *satisfied* with the Smart Angel connected medical device than those with high levels of health literacy [18,24].

Methods

Participants

We enrolled 36 participants for this study: 17 (47%) females and 19 (53%) males aged 20-64 (mean 40.75, SD 14.45) years. The inclusion criteria were that participants had to (1) have a 4G connection at home, (2) be under 70 years of age, (3) be eligible for outpatient surgery, and (4) not be at home alone. All participants were native French speakers and signed a consent form after being informed of the study's progress. The study was in line with the ethical recommendations of the Declaration of Helsinki. The participants were recruited on a voluntary basis, and no compensation was offered. Handover of the Smart Angel device took place at the participant's home or workplace.

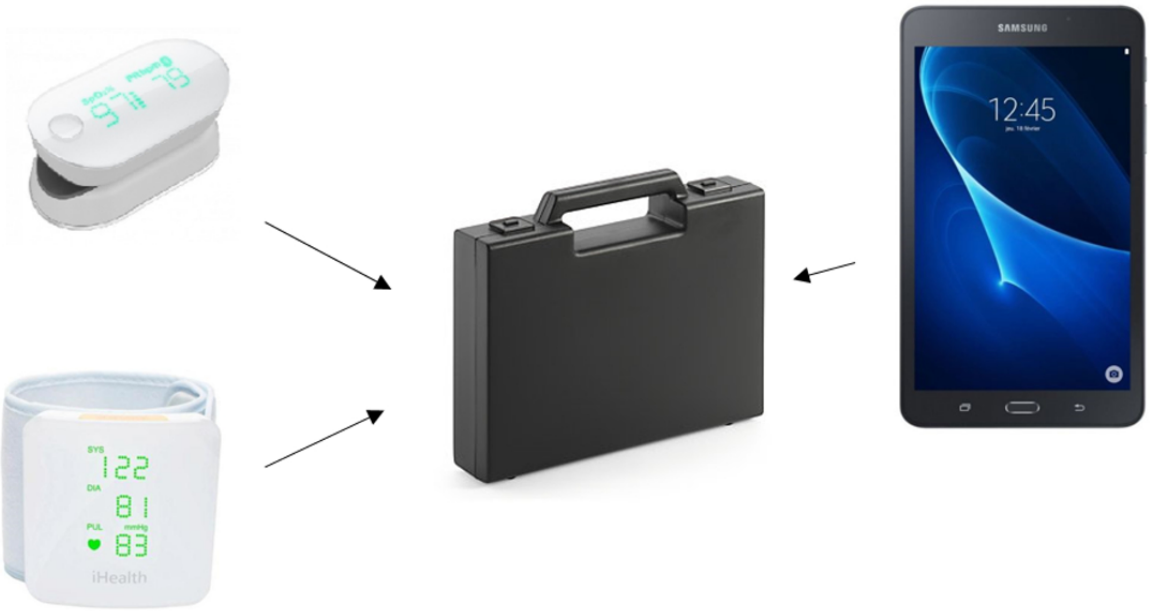
Materials and Measurements

The materials for this study included (1) the Smart Angel device, (2) personas and their scenarios, and (3) questionnaires (ie, 2 questionnaires assessing the level of health literacy, namely, the NVS and the HLS-EU; a questionnaire relating to sociodemographic data; and a questionnaire assessing satisfaction, namely, the SUS).

The Smart Angel Device

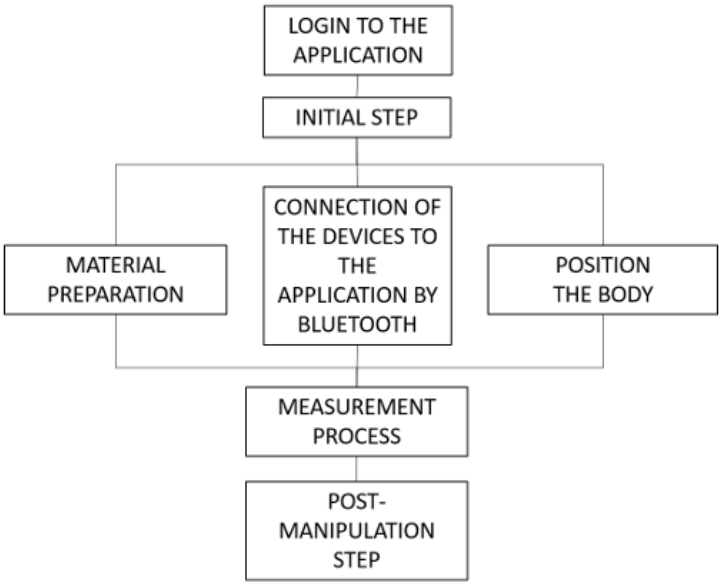
The Smart Angel device is designed by Evolucare Technologies. It consists of a Samsung 9-inch tablet with the Smart Angel application and 2 connected devices, a wrist blood pressure monitor (iHealth BP7) for blood pressure measurement and an oximeter (iHealth Oximeter PO3) for oxygen saturation and pulse measurement, which are available for the general public with European certification (Figure 1). To use the Smart Angel device, it is necessary to access the Smart Angel application and perform a digital medical “appointment” from a tablet application.

Figure 1. The Smart Angel components. Upper left: a pulse oximeter (iHealth Oximeter PO3); lower left: a wrist blood pressure monitor (iHealth BP7); right: a tablet with the Smart Angel application.



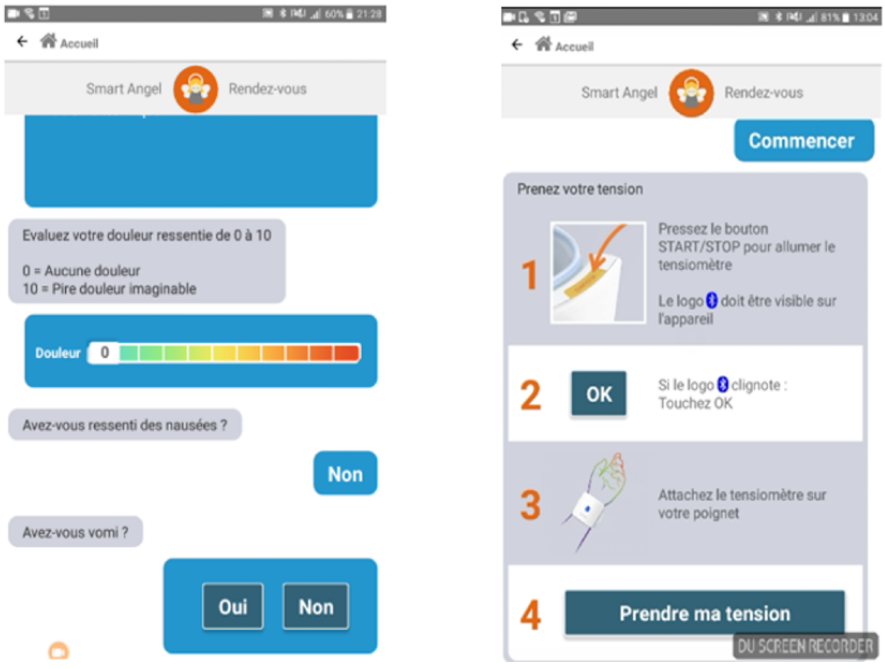
The patient is given step-by-step instructions for connecting to and taking measurements with the blood pressure monitor and the pulse oximeter. The procedures for using the blood pressure monitor and pulse oximeter were built into the application; they include text and images for each step of the operation. For the 2 connected devices, the participant must first have a correct body position to then connect the equipment, install it correctly on themselves, start the measurement, and then remove and switch off the equipment. A schematic representation of this procedure is shown in Figure 2.

Figure 2. Schematic representation of the main steps in the use of the Smart Angel device.



Once the blood pressure or oxygenation measurement has been taken, the patient's health data are displayed on a colored gauge (from green to orange) according to the level of severity of the constant collected (Figure 3). Then, the user is presented with a questionnaire with various items related to general health, pain, sleep, and nausea. These items are presented either in simple-choice question format (eg, “How are you feeling today? Good, not good, not good at all”) or on a Likert scale (eg, “Rate your pain on a scale of 1 to 10”).

Figure 3. Screenshot of the Smart Angel application (Evolucare Technologies). Left: a form presenting an overview of the subjective state of health; right: the procedure for using the monitor.



Personas and Scenarios

We constructed 5 personas and their scenarios based on statistical surveys of outpatient surgical procedure types in France [45] and observations made in the field [46]. Generally used in the design phase, the personas method draws on the theory of mind and the theory of stereotypes and can provoke certain emotional states [47]. The personas scenarios were presented to the participants as audiovisual cartoons. All scenarios were constructed in the same way. Only the type of operation and the cause of the operation changed, according to each persona. An example of a persona is presented in Multimedia Appendix 1.

Questionnaires

Measuring Health Literacy: Objective (NVS) and Subjective (HLS-EU) Assessments

Given the limited options of French-translated and validated health literacy questionnaires, we chose to use 2 health literacy questionnaires for a holistic view of this multidimensional skill:

the NVS and the 16-item Health Literacy Survey (HLS-EU-Q16):

The French-translated [48] NVS [40] is a validated test assessing a patient's ability to comprehend reading material and manipulate numbers (numeracy). Consequently, the NVS provides an objective assessment of health literacy level. Participants were asked to use an ice cream nutrition label to answer 6 questions (eg, "If I am allergic to peanuts, can I eat this ice cream?" Answer: "No, because the ice cream contains traces of peanut oil"). The total sum of the items (0-6 points) classified respondents into 3 categories: 0-1 point = inadequate health literacy; 2-3 points = problematic health literacy; 4-6 points = sufficient health literacy. The interitem reliability of the NVS in this study was good (Cronbach α=.883) [49].

The French-translated [50] HLS-EU-Q16 [31] is the short version of the HLS questionnaire. This version is composed of 16 items, 13 of which assess the 4 types of health literacy skills: the ability to access, understand, evaluate, and apply health information. Respondents were asked to rate their own ability to access information (eg, "Please rate, on a scale of very easy to very difficult, how easy is it for you to understand your doctor's or pharmacist's instructions on how to take your medication?"). Consequently, the HLS-EU-Q16 provides a subjective assessment of health literacy level. Answers are provided in 4 categories, on a 4-point Likert scale ranging from "very easy" to "very difficult." To calculate the total score, the answers "easy" and "very easy" were assigned 1 point per item, and the answers "difficult" and "very difficult" were assigned 0 points per item. The total sum of the items (0-16 points) classified respondents into 3 categories: 0-8 points = inadequate health literacy; 9-12 points = problematic health literacy; 13-16 points = sufficient health literacy. The interitem reliability of the HSL-EU in this study was good (Cronbach α=.803) [49].

Sociodemographic Measurements (Age, Education Level, Technophilia, Etc)

This questionnaire includes the following personal details: age, gender, educational level, residential area, technophilia, and hospital experience. IT experience was measured by 2 items, adapted from Agarwal and Prasad [51], related to the participant's use of and willingness to explore IT innovations (eg, "Which of these technologies do you use and how often?"). On a 5-point Likert scale, the possible answers ranged from "never" to "very often."

Measuring Usability (ISO 9241-11:2018)

Measuring Effectiveness

Effectiveness was measured by counting the number of manipulation errors, such as not putting the blood pressure cuff in the correct position. With respect to the use of the monitor, 5 categories of errors were identified: the participant (1) did not position the monitor correctly, (2) incorrectly directed the monitor toward the palm of the hand, (3) did not position the forearm correctly, (4) moved during the measurement, or (5) did not connect the monitor's Bluetooth to the tablet. Regarding the use of the pulse oximeter, 4 categories of error were identified: the participant (1) did not position the oximeter the right way, (2) did not insert the finger as far as the sensor, (3) removed the oximeter too early during the measurement, or (4) did not connect the Bluetooth from the oximeter to the tablet. With the tablet, 1 type of error was observed: the participant did not enter the appointment in the application. A scoring grid was used to identify these manipulation errors. When the participant made several attempts, we recorded the cumulative number of errors.

Measuring Efficiency

Measuring efficiency was based on the manipulation duration times of the various device tools for 3 measurements: blood pressure monitor manipulation, pulse oximeter manipulation, and total manipulation of the device, including the complete appointment. These times were measured from the time participants first touched the device (monitor, pulse oximeter, or tablet) to the time they turned it off after taking the measurement.

Measuring Satisfaction

Satisfaction was measured using the SUS. This "quick and dirty" questionnaire [52] consists of 10 items with 5 response options on a Likert scale (ranging from "strongly disagree" to "strongly agree"), which allows for a subjective assessment of usability [53]. We used an adapted and validated version [54], in which we replaced the term "system" with the term "medical device." Scores were calculated according to the recommendations of Brooke [52] and ranged from 0 to 100. Lower scores indicate low usability.

Procedure

The average duration of this experiment was 45 minutes. The selected participants did not come out of ambulatory surgery. Participants were first invited to choose among 5 proposed personas to allow them to project themselves into the needs of future users of the Smart Angel device [55]. The persona chosen had to be consistent with at least the participant's age, profession, and previous surgery. Then, the researcher demonstrated the use of the Smart Angel device to the participant for about 3 minutes, sharing information about the correct manipulation of the device (eg, "The monitor should always be at heart level"). Participants were asked to complete 3 questionnaires: the sociodemographic data questionnaire, the HLS-EU-Q16, and the NVS. Then they were asked to operate the Smart Angel device by taking a blood pressure measurement followed by an oxygen saturation measurement, and finally, by completing the general health questionnaire. There was no time limit for this. The participants were filmed during the process. The researcher could only intervene in the event of a technical problem (eg, battery problem). Finally, after the experiment, the participant had to respond to the SUS.

Data Analysis

The videos were analyzed using BORIS (Behavioral Observation Research Interactive Software) [56], which collected data on effectiveness and efficiency. Results were analyzed using SPSS software (version 22; IBM Corp). Each user characteristic was systematically compared to usability components, including effectiveness, efficiency, and satisfaction. For the health literacy measurement, we first analyzed the HLS-EU-Q16 result and then the NVS result. Bivariate correlations, ANOVAs, and

Student *t* tests were performed when the sample met the homoscedasticity criteria, while nonparametric tests (Kruskal-Wallis and Mann-Whitney) were performed when the sample did not meet these criteria.

Interjudge Reliability: Objective Measures of Effectiveness and Efficiency

We used intraclass correlation (ICC) to verify interjudge reliability for quantitative data [57]. A 33% double coding of the collected video data was performed. The mean ICC measurement for total manipulation time (efficiency) was 0.978, with a 95% confidence interval of 0.918 to 0.994 ($F_{11,11}$=45.436; *P*<.001). The mean ICC measurement (efficiency) for manipulating the monitor was 0.988, with a 95% confidence interval of 0.954 to 0.997 ($F_{11,11}$=81.635; *P*<.001). The mean ICC measurement (efficiency) for manipulating the pulse oximeter was 0.956, with a 95% confidence interval of 0.838 to 0.988 ($F_{11,11}$=22.955; *P*<.001). The mean ICC measurement (efficiency) for manipulating the tablet was 0.906, with a 95% confidence interval 0.652 to 0.975 ($F_{11,11}$=10.688; *P*<.001). The mean measure of the number of manipulation errors (effectiveness) was 0.952, with a 95% confidence interval of 0.842 and 0.985 ($F_{11,11}$=20.789; *P*<.001).

Results

Effects of User Characteristics on Usability

The correlations between user characteristics and usability components (ie, effectiveness = number of manipulation errors; efficiency = manipulation time; satisfaction = SUS score) were systematically analyzed (Table 1).

Table 1. Descriptive analyses of user characteristics, user experiences in health, medical devices, and technology (n=36).

Characteristics	Value	Average effectiveness, number of errors (SD)	Average efficiency, manipulation time in seconds (SD)	Average satisfaction, SUS score (SD)
Age in years, mean (SD)	40.75 (14.45)	N/A[a]	N/A	N/A
Gender, n (%)				
Male	19 (52.8)	1.21 (1.27)	362.09 (144.16)	87.24 (11.18)
Female	17 (47.2)	2.06 (1.25)	373.91 (126.3)	81.03 (11.73)
Education, n (%)				
Secondary education	5 (13.9)	2.8 (1.64)	337.96 (89.67)	77 (9.75)
Higher education, 1st cycle	11 (30.6)	1.36 (1.1)	334.99 (106.25)	87.05 (13.82)
Higher education, 2nd cycle	11 (30.6)	1.64 (1.2)	412.95 (198.6)	82.73 (13.34)
Higher education, 3rd cycle	9 (25)	1.22 (1.3)	368.77 (80.88)	86.94 (5.97)
Residential area, n (%)				
Rural	6 (16.7)	1 (0.89)	362.92 (76.48)	88.75 (6.85)
Semi-urban	5 (13.9)	1.8 (2.05)	339.11 (72.3)	87 (11.37)
Urban	25 (64.9)	1.72 (1.24)	374.52 (154.78)	82.7 (12.62)
Persona chosen, n (%)				
Persona 1	8 (22.2)	*[b]	*	*
Persona 2	8 (22.2)	*	*	*
Persona 3	8 (22.2)	*	*	*
Personal 4	4 (11.1)	*	*	*
Persona 5	8 (22.2)	*	*	*
Health care experience with operations, n (%)				
Yes	32 (88.9)	1.59 (1.21)	376.18 (136.24)	85,39 (11.72)
No	4 (11.1)	1.75 (2.22)	299.55 (106.9)	75.62 (8)
Health care experience with outpatient operations, n (%)				
Yes	18 (50)	1.39 (1.33)	367.15 (142.09)	86.11 (11.8)
No	18 (50)	1.83 (1.29)	368.19 (130)	82,5 (11.66)
Health care experience with suffering from a chronic illness, n (%)				
Yes	11 (30.6)	1.27 (1.35)	386.68 (184.62)	81.14 (15.26)
No	25 (69.4)	1.76 (1.3)	359.3 (108.77)	85.7 (9.8)
Medical device experience with taking blood pressure, n (%)				
Yes	24 (66.7)	1.54 (1.32)	361.43 (131.26)	63.3 (11.22)
No	12 (33.3)	1.75 (1.36)	380.15 (145.06)	86.25 (12.9)
Medical device experience with blood oxygenation testing, n (%)				
Yes	5 (13.9)	0.4 (0.55)	309.16 (70.87)	89 (8.02)
No	31 (86.1)	1.81(1.3)	377.11 (140.31)	83.56 (12.12)
Information technology experience with ease of use of tablet/computer/telephone, n (%)				
Very comfortable	23 (63.9)	1.35 (1.23)	360.16 (120.21)	86.85 (11.24)
Relatively comfortable	11 (30.6)	2.27 (1.42)	401.34 (166.78)	78.18 (11.78)
Moderately comfortable	2 (5.6)	1 (0)	268.81 (33.95)	88.75 (5.3)
Rather uncomfortable	0 (0)	__[c]	—	—

Characteristics	Value	Average effectiveness, number of errors (SD)	Average efficiency, manipulation time in seconds (SD)	Average satisfaction, SUS score (SD)
Not at all comfortable	0 (0)	—	—	—
Frequency of use of technology, n (%)				
Very often (every day)	5 (13.9)	1 (1.22)	314.42 (48.65)	92.5 (6.85)
Often (several times a week)	12 (33.3)	1.25 (1.36)	364.46 (133.26)	90.42 (6.47)
Rarely (from time to time)	17 (47.2)	2 (1.27)	364.08 (99.16)	79.26 (11.38)
Very rarely (occasionally)	1 (2.8)	3 (—)	856.33 (—)	55 (—)
Never	1 (2.8)	1 (—)	244.8 (—)	85 (—)

[a]N/A: not applicable.

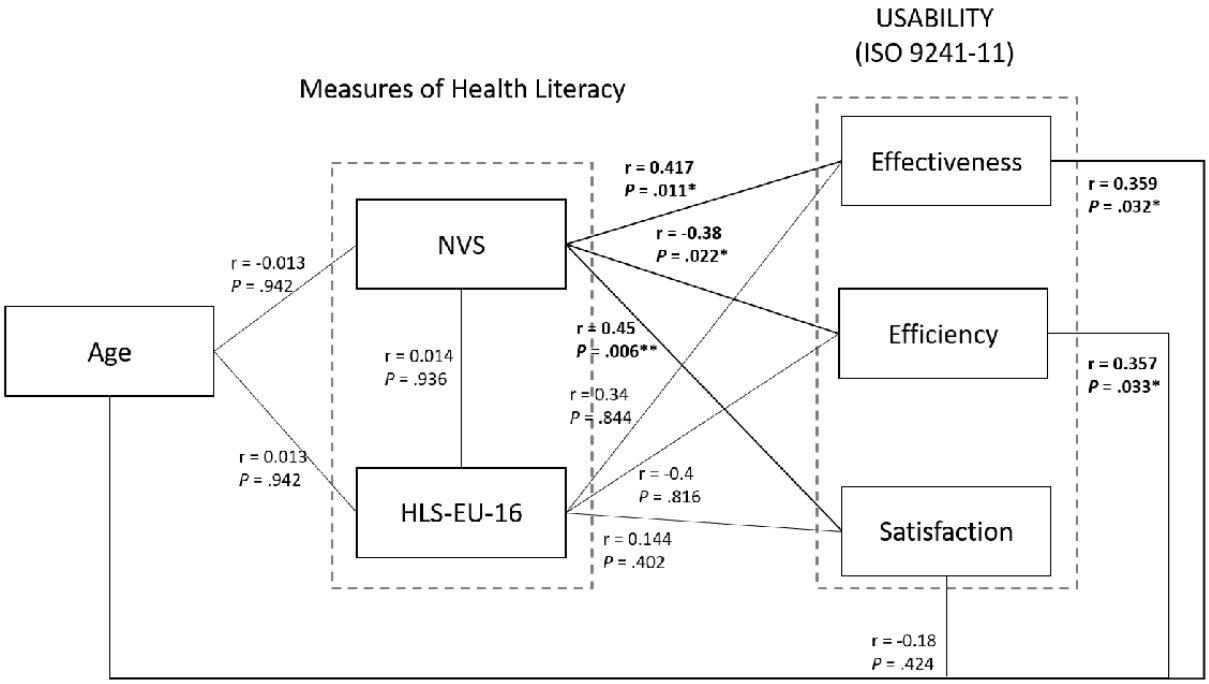
[b]*Highly correlated to the ages of the participants.

[c]— Not available.

Age

The age of the participants (mean 40.75, SD 14.45, range 20-64 years) is significantly correlated (positively and weakly) with the number of manipulation errors (effectiveness: r=0.359; *P*=.03) and manipulation time (efficiency: r=0.357; *P*=.03). On the other hand, there was no significant correlation between age and SUS score (satisfaction: r=-0.138; *P*=.42). In addition, it is important to note that age is not correlated with the literacy level of the HLS-EU-Q16 (r=0.013; *P*=.94) or the NVS (r=-0.013; *P*=.94; Figure 4).

Figure 4. Schematic representation of correlations between age, measurements of participants' health literacy, and usability (ISO 9241-11). HLS-EU-Q16: 16-item European Health Literacy Scale; NVS: Newest Vital Sign. *P<0.5; **P<0.01.



Technophilia

The IT experience of participants had no impact on the number of manipulation errors (effectiveness: $F_{5,30}$=1.229; *P*=.32) or manipulation time (efficiency: $F_{5,30}$=1.39; *P*=.26). On the other hand, there was a significant correlation between IT experience and SUS score (satisfaction: $\chi(3)$=8.671; *P*=.03).

Moreover, previous experience of using medical devices that allow users to take their own blood pressure did not influence the number of manipulation errors (effectiveness: t_{34}=0.443; *P*=.66), the manipulation time (efficiency: t_{34}=0.39; *P*=.55), or the SUS score (satisfaction: Mann-Whitney U=104; *P*=.19). Previous experience of using medical devices for taking oxygen levels had a significant effect on the number of manipulation errors (effectiveness: t_{34}=2.359; *P*=.02; $\eta 2$=0.14), but this effect was not significant on the manipulation time (efficiency: t_{34}=1.052; *P*=.30) or the SUS score (satisfaction: t_{34}=-0.965; *P*=.34).

Educational Level

Educational level had no impact on usability in terms of the number of manipulation errors (effectiveness: $F_{3,32}$=1.889; *P*=.15), manipulation time (efficiency: $F_{3,32}$=0.698; *P*=.56), and SUS score (satisfaction: $F_{3,32}$=1.076; *P*=.37).

Health Literacy

Systematic analyses were performed comparing the level of literacy (HLS-EU-Q16 and NVS) with each of the components of usability (effectiveness, efficiency, and satisfaction, as per the ISO 9241-11, 2018); Table 2 presents a descriptive representation of the results of the 2 health literacy questionnaires.

Table 2. Descriptive statistics of the 16-item European Health Literacy Survey (HLS-EU-Q16) and the New Vital Sign (NVS) questionnaires.

Statistic	HLS-EU-Q16	NVS
Mean (SD), range	12.97 (2.952), 5-16	4.17 (2.223), 0-6
Inadequate health literacy, n (%)	3 (8.3)	6 (16.7)
Problematic health literacy, n (%)	9 (25)	7 (19.4)
Sufficient health literacy, n (%)	24 (66.7)	23 (63.9)

HLS-EU-Q16 Questionnaire Results

There was no significant correlation between the results of the HLS-EU-Q16 and usability, either in terms of the number of manipulation errors (effectiveness: r=0.34; *P*=.84), manipulation time (efficiency: r=-0.40; *P*=.82), or the SUS score (satisfaction: r=0.144; *P*=.40). After correlation analysis, participants were clustered according to the HLS-EU-Q16 measures (Table 2), following the recommendations of Sørensen et al [31]. No intergroup differences could be observed between the HLS-EU-Q16 results and usability (Table 3) in terms of the number of manipulation errors (effectiveness: $F_{2.33}$=0.277; *P*=.76), manipulation time (efficiency: $F_{2.33}$=0.015; *P*=.99), and the SUS score (satisfaction: $F_{2.33}$=0.483; *P*=.62).

Table 3. Analyses of the 16-item European Health Literacy Survey (HLS-EU-Q16) score according to usability (effectiveness, efficiency, and satisfaction; n=36).

HLS-EU-Q16 score classification group (n=36)	Effectiveness[a], average number of errors (SD)	Efficiency[b], average manipulation time (SD)	Satisfaction[c], SUS[d] score (SD)
Inadequate health literacy (n=3)	1.67 (2.08)	373.26 (88.76)	83.33 (3.82)
Problematic health literacy (n=9)	1.89 (1.27)	373.35 (98.53)	81.11 (14.53)
Sufficient health literacy (n=24)	1.50 (1.28)	364.84 (152.73)	85.62(11.3)

[a]ANOVA: $F_{2,33}$=0.277; *P*=.76.

[b]ANOVA: $F_{2,33}$=0.015; *P*=.99.

[c]ANOVA: $F_{2,33}$=0.483; *P*=.62.

[d]SUS: System Usability Survey.

NVS Questionnaire Results

There was a significant mean-size correlation between the results of the French version of the NVS questionnaire and usability (Table 4) in terms of the number of manipulation errors (effectiveness: r=-0.417; *P*=.01), manipulation time (efficiency: r=-0.38; *P*=.02), and the SUS score (satisfaction: r=0.45; *P*=.006). In other words, the higher a participant's level of health literacy (measured using NVS), the fewer manipulation errors they made (ie, they are more effective), the faster they manipulate (ie, they are more efficient), and the higher their SUS score will be (ie, they will be more satisfied).

After analyzing the correlations, the participants were clustered according to the NVS measurements (Table 2), following recommendations [40]. No intergroup differences could be observed between NVS literacy and usability (Table 4) except for the number of errors (effectiveness: $\chi 2$=6.679; *P*=.04).

Further intergroup analysis (Figure 4) shows a significant effect between the inadequate-health-literacy and sufficient-health-literacy groups as a function of the number of manipulation errors (effectiveness: Mann-Whitney U=27; *P*=.02).

Table 4. Analyses of the New Vital Sign (NVS) results according to usability (effectiveness, efficiency, and satisfaction; n=36).

NVS score classification group (n=36)	Effectiveness[a], average number of errors (SD)	Efficiency[b], average manipulation time (SD)	Satisfaction[c], SUS[d] score (SD)
Inadequate health literacy (n=6)	2.67 (0.816)	463 (165.18)	77.08 (14.27)
Problematic health literacy (n=7)	1.71 (0.756)	387.72 (219.2)	80.71 (15.05)
Sufficient health literacy (n=23)	1.30 (1.43)	336.7 (75.79)	87.28 (9.07)

[a]Kruskal-Wallis test: χ^2=6.679; *P*=.035, where *P*<.05 is significant.

[b]Kruskal-Wallis test: χ^2=3.07; *P*=.21.

[c]ANOVA: $F_{2,33}$=2.392, *P*=.11; Kruskal-Wallis test: χ^2=2.618, *P*=.27.

[d]SUS: System Usability Survey.

Discussion

Principal Findings

This study's objective was to better understand the relationships between 4 user characteristics (age, education, technophilia, and health literacy) and usability [3] (defined here as effectiveness, efficiency, and satisfaction) with regard to the use of the Smart Angel device. To do this, sociodemographic data were collected, literacy levels were investigated using the HLS-EU-Q16 [31] and the NVS [40], and usability measures were performed (errors and manipulation time, and SUS questionnaire).

We made 4 hypotheses that age (H1), technophilia (H2), and health literacy (H4) would have an impact on usability, while education level (H3) would not. Our first hypothesis (H1) was that older users would be less effective, efficient, and satisfied with the device compared to younger users. We can partially validate this hypothesis. The results show that the younger the individuals are, the less likely they are to make manipulation errors (ie, they are more effective) and the faster they manipulate the device (ie, they are more efficient). On the other hand, we did not observe any difference between the age of the subjects and the SUS score (satisfaction). All these results are in line with previous research [19,20,25,26]. Indeed, younger users are more effective (eg, Jones and Caird's glucometer [25]) and efficient (eg, Mykityshyn et al's glucometer [26] and Van der Vaart et al's application for narcoleptics [20]) compared to older users, with a positive and medium correlation [20]. However, younger users are as satisfied (SUS score) with the device as older users, which is consistent with the findings of Liang et al [19] while at variance with those of Georgsson and Staggers [11].

Our second hypothesis (H2) focused on technophilia (experience of information technology and medical devices). The results provide partial validation of this hypothesis, as no correlation was observed between IT experience and usability in terms of effectiveness and efficiency. On the other hand, the technophile participants had a significantly better SUS score (satisfaction) than participants with a low level of technophilia. While these results are consistent with those of Harte et al [27], they contradict previous works [11,23]. We explain these results by a relatively homogeneous representation of IT experience as a function of the age of participants in our sample. We believe that these items [51] highlight the subjective representation of technology use (in relation to age) rather than actual performance in the use of hardware. It is possible that older people may feel that they can properly manipulate a tablet without using other features available in the tool. They would then consider themselves to be quite technophilic, as they would be effective in the day-to-day use of the technology. However, their real capacity to adapt to the technologies is unknown. For example, if an update were to be performed on one of the applications commonly used, it is possible that this would destabilize the manipulation carried out by these individuals.

We also observed a correlation between experience with medical devices and usability. However, previous experience in the use of a blood pressure monitor had no impact on usability. Conversely, previous experience in the use of a pulse oximeter had a significant effect on effectiveness. Participants who had previously manipulated a pulse oximeter made significantly fewer errors than those who had never manipulated a pulse oximeter. In contrast, previous experience using a pulse oximeter had no effect on efficiency and satisfaction. All subjects who reported previous use of a pulse oximeter also reported previous manipulation of a blood pressure monitor. This result suggests that prior use of a pulse oximeter in combination with a blood pressure monitor would facilitate manipulation of the Smart Angel device in terms of effectiveness. We believe that participants who are accustomed to using this type of complex device are accustomed to being involved in health issues, which may be evidence of strong patient involvement in their own health [58].

Our third hypothesis (H3) was concerned with the lack of correlation between education level and usability. The results supported our hypothesis, as no significant correlation was found between participants' level of education and usability in terms of effectiveness, efficiency, and satisfaction. These results are also consistent with previous works [11,19,20].

Finally, the fourth hypothesis (H4) postulated that health literacy influences usability (effectiveness, efficiency, and satisfaction). The HLS-EU-Q16 scores showed no effect on usability (Figure 4). In contrast, the NVS scores showed a significant effect on the number of manipulation errors (effectiveness), manipulation time (efficiency), and SUS score (satisfaction). This is consistent with the results of previous studies [18,28,29]. Significant and medium-sized correlations between the NVS score and each of

the usability dimensions were observed (Figure 4). This suggests that the higher the literacy level of the participants, the fewer manipulation errors they make (ie, the more effective they are), the faster they are (ie, the more efficient they are), and the higher the SUS score will be (ie, the more satisfied they are). However, after clustering the participants as recommended [40], there is a significant correlation between NVS literacy level and the number of errors (effectiveness) but no correlation with the manipulation time (efficiency) and the SUS score (satisfaction). Participants with a sufficient literacy level made significantly fewer errors than those with inadequate or problematic literacy.

It is important to note that the HLS-EU and NVS results are contradictory and demonstrate the complexity of health literacy assessment. In addition, our results suggest that the HLS-EU questioning the participants' subjective abilities to access health information and make decisions introduces a significant bias in the measurement of health literacy. Some participants may claim to have no difficulty using health information, but there is no verification that this is, in fact, the case. Conversely, the NVS instrument appears to be better suited to gathering information on subjects' cognitive abilities, as it is a test that collects information on participants' thought processes when reading a food label, thus providing a more objective assessment of health literacy.

Conclusions and Research Prospects

This study provides theoretical insight into the effects of user characteristics (eg, age, experience, education, and health literacy) through the use of personas with respect to usability (effectiveness, efficiency, and satisfaction, according to ISO 9241-11 [3]) in the case of the Smart Angel connected medical device. This study provides a methodological contribution insofar as it revealed the differences in data collection between the NVS and the HLS-EU-Q16, thus demonstrating the importance of continuing research in the field of health literacy measurement tools. In addition, these results allow us to better understand the importance of the impact of technophilia among older people with a sufficient level of health literacy for usability.

The results of this study suggest 4 research prospects. First, the relevance of the personas method in the prototype evaluation phase has never been proven. This method is classically used in the design phase by designers (ergonomists, designers, engineers, and even future users) but more rarely used in an evaluation framework. To validate this method in this new context of use in the evaluation phase, it would be necessary to reproduce this study by adding a control group (ie, a group for whom the personas are not presented). Secondly, the training carried out by the researcher could be adapted according to the literacy levels of the participants. Indeed, the main difficulty in the use of a medical device is understanding the procedures, and this cannot be achieved if there is insufficient upstream training [59]. Training should certainly be adapted to the ages and literacy levels of the participants. Demonstration by the researcher may be sufficient for groups with adequate levels of health literacy. Conversely, for groups with inadequate or problematic levels of health literacy, further instruction should be considered. Third, the choice of questionnaire is a crucial step in measuring health literacy. Indeed, we observed a significant disparity in results between the HLS-EU-Q16 and the NVS. As already discussed, these 2 questionnaires do not appear to assess the same dimensions of health literacy. Further work is needed to understand what exactly is being assessed by each of the health literacy questionnaires. We believe that it is better to evaluate this skill with objective assessments. In the same way, it would have been interesting to perform objective measurements of technophilia.

Finally, beyond health literacy, it would now be appropriate to measure the level of eHealth literacy [20]. Unfortunately, there is no valid questionnaire in French on this subject. Thus, more systematic translations and adaptations of these tools should be considered in future studies.

Currently, as a result of this study, the Smart Angel device is in clinical trials where usability tests continue to be carried out in in situ conditions.

Acknowledgments

This project was financially supported by the European Regional Development Fund (ERDF), Evolucare, and the Programme d'Investissements d'Avenir (PIA), which is one of 5 research and development projects structuring competitiveness of the CSPP.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of a persona.
[PDF File (Adobe PDF File), 707 KB - humanfactors_v8i1e24846_app1.pdf]

References

1. Ghosh S, Sallam S. Patient satisfaction and postoperative demands on hospital and community services after day surgery. Br J Surg 1994 Nov;81(11):1635-1638. [doi: 10.1002/bjs.1800811124] [Medline: 7677820]

2. International Organization for Standardization (ISO). Medical devices - Application of risk management to medical devices. 2019. URL: https://www.iso.org/standard/72704.html [accessed 2019-03-03] [WebCite Cache ID https://www.iso.org/obp/ui/#iso:std:72704:en]
3. International Organization for Standardization (ISO). Ergonomics of human-system interaction - Part 11: Usability: Definitions concepts. 2018. URL: https://www.iso.org/standard/63500.html [accessed 2018-03-03] [WebCite Cache ID https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en]
4. Carayon P, Hoonakker P. Human Factors and Usability for Health Information Technology: Old and New Challenges. Yearb Med Inform 2019 Aug;28(1):71-77 [FREE Full text] [doi: 10.1055/s-0039-1677907] [Medline: 31419818]
5. Cifter AS. Blood Pressure Monitor Usability Problems Detected Through Human Factors Evaluation. Ergonomics in Design 2017 May 22;25(3):11-19. [doi: 10.1177/1064804617691397]
6. Greysen SR, Chin Garcia C, Sudore RL, Cenzer IS, Covinsky KE. Functional impairment and Internet use among older adults: implications for meaningful use of patient portals. JAMA Intern Med 2014 Jul 01;174(7):1188-1190 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1864] [Medline: 24839165]
7. Kortum P, Peres SC. Evaluation of Home Health Care Devices: Remote Usability Assessment. JMIR Hum Factors 2015 Jun 05;2(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.4570] [Medline: 27025664]
8. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
9. 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]
10. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
11. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
12. Jaspers MW. A comparison of usability methods for testing interactive health technologies: methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
13. Nelson LA, Mulvaney SA, Johnson KB, Osborn CY. mHealth Intervention Elements and User Characteristics Determine Utility: A Mixed-Methods Analysis. Diabetes Technol Ther 2017 Jan;19(1):9-17 [FREE Full text] [doi: 10.1089/dia.2016.0294] [Medline: 28099052]
14. Peute LW, Driest KF, Marcilly R, Bras Da Costa S, Beuscart-Zephir M, Jaspers MWM. A framework for reporting on human factor/usability studies of health information technologies. Stud Health Technol Inform 2013;194:54-60. [Medline: 23941930]
15. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
16. Borsci S, Uchegbu I, Buckle P, Ni Z, Walne S, Hanna GB. Designing medical technology for resilience: integrating health economics and human factors approaches. Expert Rev Med Devices 2018 Jan 20;15(1):15-26. [doi: 10.1080/17434440.2018.1418661] [Medline: 29243500]
17. Grebin SZ, Echeveste MES, Magnago PF, Tanure RLZ, Pulgati FH. [Analytical strategy for assessment of usability of medical devices from the user's perspective: a study of hemodialysis patients]. Cad Saude Publica 2018 Aug 20;34(8):e00074417 [FREE Full text] [doi: 10.1590/0102-311X00074417] [Medline: 30133654]
18. Jensen JD, King AJ, Davis LA, Guntzviller LM. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. J Aging Health 2010 Sep;22(6):804-826. [doi: 10.1177/0898264310366161] [Medline: 20495159]
19. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability Study of Mainstream Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation. JMIR Mhealth Uhealth 2018 Nov 08;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]
20. van der Vaart R, van Driel D, Pronk K, Paulussen S, Te Boekhorst S, Rosmalen JGM, et al. The Role of Age, Education, and Digital Health Literacy in the Usability of Internet-Based Cognitive Behavioral Therapy for Chronic Pain: Mixed Methods Study. JMIR Form Res 2019 Nov 21;3(4):e12883 [FREE Full text] [doi: 10.2196/12883] [Medline: 31750839]
21. Monkman H, Kushniruk A. A health literacy and usability heuristic evaluation of a mobile consumer health application. Stud Health Technol Inform 2013;192:724-728. [Medline: 23920652]
22. Sonderegger A, Schmutz S, Sauer J. The influence of age in usability testing. Appl Ergon 2016 Jan;52:291-300. [doi: 10.1016/j.apergo.2015.06.012] [Medline: 26360221]
23. Sparkes J, Valaitis R, McKibbon A. A usability study of patients setting up a cardiac event loop recorder and BlackBerry gateway for remote monitoring at home. Telemed J E Health 2012;18(6):484-490. [doi: 10.1089/tmj.2011.0230] [Medline: 22676379]

24. 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;36(1-2):45-60 [FREE Full text] [doi: 10.1016/s1532-0464(03)00056-x] [Medline: 14552846]
25. Jones J, Caird JK. The Usability of Blood Glucose Meters: Task Performance Differences Between Younger and Older Age Groups. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2017 Sep 28 Presented at: Proceedings of the Human Factors and Ergonomics Society Annual Meeting Internet; 2017 Sep 1; 61st International Annual Meeting, JW Marriott, Austin, Texas, USA p. 604-608. [doi: 10.1177/1541931213601636]
26. Mykityshyn AL, Fisk AD, Rogers WA. Learning to use a home medical device: mediating age-related differences with training. Hum Factors 2002;44(3):354-364. [doi: 10.1518/0018720024497727] [Medline: 12502154]
27. Harte R, Hall T, Glynn L, Rodríguez-Molinero A, Scharf T, Quinlan LR, et al. Enhancing Home Health Mobile Phone App Usability Through General Smartphone Training: Usability and Learnability Case Study. JMIR Hum Factors 2018 Apr 26;5(2):e18 [FREE Full text] [doi: 10.2196/humanfactors.7718] [Medline: 29699969]
28. Czaja SJ, Zarcadoolas C, Vaughon WL, Lee CC, Rockoff ML, Levy J. The usability of electronic personal health record systems for an underserved adult population. Hum Factors 2015 May 05;57(3):491-506 [FREE Full text] [doi: 10.1177/0018720814549238] [Medline: 25875437]
29. 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]
30. Chaniaud N, Métayer N, Megalakaki O, Loup-Escande E. Effect of Prior Health Knowledge on the Usability of Two Home Medical Devices: Usability Study. JMIR Mhealth Uhealth 2020 Sep 21;8(9):e17983 [FREE Full text] [doi: 10.2196/17983] [Medline: 32955454]
31. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12(1):80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
32. Williams MV, Baker DW, Parker RM, Nurss JR. Relationship of functional health literacy to patients' knowledge of their chronic disease. A study of patients with hypertension and diabetes. Arch Intern Med 1998 Jan 26;158(2):166-172. [doi: 10.1001/archinte.158.2.166] [Medline: 9448555]
33. Mõttus R, Johnson W, Murray C, Wolf MS, Starr JM, Deary IJ. Towards understanding the links between health literacy and physical health. Health Psychol 2014 Feb;33(2):164-173. [doi: 10.1037/a0031439] [Medline: 23437854]
34. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
35. Duteille F. Compétences en santé déficientes : obstacle à une prise en charge optimale. Revue Médicale Suisse, 8 2019 Apr:1016-1021 [FREE Full text] [doi: 10.1016/j.refrac.2019.04.003]
36. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol 2011 Apr;64(4):366-379. [doi: 10.1016/j.jclinepi.2010.04.005] [Medline: 20638235]
37. Haun JN, Valerio MA, McCormack LA, Sørensen K, Paasche-Orlow MK. Health literacy measurement: an inventory and descriptive summary of 51 instruments. J Health Commun 2014 Oct 14;19 Suppl 2(sup2):302-333. [doi: 10.1080/10810730.2014.936571] [Medline: 25315600]
38. Parker R. Health literacy: a challenge for American patients and their health care providers. Health Promot Int 2000;15(4):277-283 [FREE Full text] [doi: 10.1093/heapro/15.4.277]
39. Murphy P, Davis T, Long S, Jackson R, Decker B. Rapid Estimate of Adult Literacy in Medicine (REALM): A Quick Reading Test for Patients. Journal of Reading 1993;37(2):124-130 [FREE Full text]
40. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: the newest vital sign. Ann Fam Med 2005;3(6):514-522 [FREE Full text] [doi: 10.1370/afm.405] [Medline: 16338915]
41. Connor M, Mantwill S, Schulz PJ. Functional health literacy in Switzerland--validation of a German, Italian, and French health literacy test. Patient Educ Couns 2013 Jan;90(1):12-17. [doi: 10.1016/j.pec.2012.08.018] [Medline: 23089240]
42. Housten AJ, Lowenstein LM, Hoover DS, Leal VB, Kamath GR, Volk RJ. Limitations of the S-TOFHLA in measuring poor numeracy: a cross-sectional study. BMC Public Health 2018 Mar 27;18(1):405 [FREE Full text] [doi: 10.1186/s12889-018-5333-9] [Medline: 29587709]
43. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns 1999 Sep;38(1):33-42. [doi: 10.1016/s0738-3991(98)00116-5] [Medline: 14528569]
44. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). Eur J Public Health 2015 Dec;25(6):1053-1058 [FREE Full text] [doi: 10.1093/eurpub/ckv043] [Medline: 25843827]
45. Agence Technique De L'Information sure L'Hospitalisation (ATIH). Etat des lieux 2015 sur l'activité de chirurgie ambulatoire. 2016. URL: https://www.atih.sante.fr/sites/default/files/public/content/1504/note_gdr_chirambu_2015.pdf [accessed 2016-03-03] [WebCite Cache ID https://www.atih.sante.fr/sites/default/files/public/content/1504/note_gdr_chirambu_2015.pdf]

46. Loup-Escande E, Valentin A, Metayer N, Ricquebourd V, Bidard A, Damay P, et al. Adaptation de la démarche ergonomique en phase amont de la conception de technologies e-santé?xemple du projet Smart Angel. 2019 Presented at: EPIQUE; 9 au 12 juillet; Lyon p. e.
47. Lallemand C, Gronier G. Methodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs. In: Methodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs. Paris: Editions Eyrolles; 2018:712.
48. Mansfield ED, Wahba R, Gillis DE, Weiss BD, L'Abbé M. Canadian adaptation of the Newest Vital Sign©, a health literacy assessment tool. Public Health Nutr 2018 Aug;21(11):2038-2045. [doi: 10.1017/S1368980018000253] [Medline: 29692290]
49. Nunnaly J. Psychometric Theory. Chicago: McGraw Hill; 1978.
50. Rouquette A, Nadot T, Labitrie P, Van den Broucke S, Mancini J, Rigal L, et al. Validity and measurement invariance across sex, age, and education level of the French short versions of the European Health Literacy Survey Questionnaire. PLoS One 2018 Dec 6;13(12):e0208091 [FREE Full text] [doi: 10.1371/journal.pone.0208091] [Medline: 30521552]
51. Agarwal R, Prasad J. A Conceptual and Operational Definition of Personal Innovativeness in the Domain of Information Technology. Information Systems Research 1998 Jun;9(2):204-215. [doi: 10.1287/isre.9.2.204]
52. Brooke J. SUS - A quick and dirty usability scale. In: Jordan P, Thomas B, McClelland I, Weerdmeester B, editors. Usability evaluation in industry. London: CRC Press; 1996:252.
53. Lewis J, Sauro J. The Factor Structure of the System Usability Scale. In: Human Centered Design Internet Springer, Berlin, Heidelberg. 2009 Presented at: Interantion Conference on Human Centered Design; 2009; Berlin p. 94-103. [doi: 10.1007/978-3-642-02806-9_12]
54. Bangor A, Kortum PT, Miller JT. An Empirical Evaluation of the System Usability Scale. International Journal of Human-Computer Interaction 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
55. Bornet C, Brangier E. La méthode des personas: principes, intérêts et limites. Bulletin de psychologie Internet 2013 May 3;524:115-134 [FREE Full text] [doi: 10.3917/bupsy.524.0115]
56. Friard O, Gamba M. : a free, versatile open‐source event‐logging software for video/audio coding and live observations. Methods Ecol Evol 2016 May 28;7(11):1325-1330. [doi: 10.1111/2041-210x.12584]
57. Elie C, Colombet I. Methods of evaluation of reproducibility. Sang thrombose vaisseaux---Sang Thrombose Vaisseaux Internet 2011 Mar 2011;23(3):138-145. [doi: 10.1684/stv.2020.1138]
58. 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 30;27(5):520-526 [FREE Full text] [doi: 10.1007/s11606-011-1931-2] [Medline: 22127797]
59. Larsen T. Nurses' instruction of patients in the use of INR-monitors for self-management of cardio-vascular conditions: Missed instructional opportunities. Patient Educ Couns 2017 Apr;100(4):673-681. [doi: 10.1016/j.pec.2016.10.001] [Medline: 27839890]

Abbreviations

HLS-EU: European Health Literacy Survey
HLS-EU-Q16: 16-item European Health Literacy Survey
ICC: intraclass correlation
IT: information technology
NVS: Newest Vital Sign
REALM: Rapid Estimate of Adult Literacy in Medicine
SUS: System Usability Survey
S-TOFHLA: Short Test of Functional Health Literacy in Adults
TOFHLA: Test of Functional Health Literacy in Adults

Edited by A Kushniruk; submitted 07.10.20; peer-reviewed by D Michelet, D Chrimes; comments to author 23.12.20; revised version received 28.12.20; accepted 03.01.21; published 17.03.21.

Please cite as:
Chaniaud N, Megalakaki O, Capo S, Loup-Escande E
Effects of User Characteristics on the Usability of a Home-Connected Medical Device (Smart Angel) for Ambulatory Monitoring: Usability Study
JMIR Hum Factors 2021;8(1):e24846
URL: https://humanfactors.jmir.org/2021/1/e24846
doi:10.2196/24846
PMID:33729161

Original Paper

Emotional Reactions and Likelihood of Response to Questions Designed for a Mental Health Chatbot Among Adolescents: Experimental Study

Audrey Mariamo[1], BA; Caroline Elizabeth Temcheff[1], PhD; Pierre-Majorique Léger[2], PhD; Sylvain Senecal[3], PhD; Marianne Alexandra Lau[1], BA

[1]Department of Educational and Counselling Psychology, McGill University, Montreal, QC, Canada

[2]Department of Information Technologies, HEC Montreal, Montreal, QC, Canada

[3]Department of Marketing, HEC Montreal, Montreal, QC, Canada

Corresponding Author:
Audrey Mariamo, BA
Department of Educational and Counselling Psychology
McGill University
3700 McTavish St
Montreal, QC, H3A 1Y2
Canada
Phone: 1 5149697302
Email: audrey.mariamo@mail.mcgill.ca

Abstract

Background: Psychological distress increases across adolescence and has been associated with several important health outcomes with consequences that can extend into adulthood. One type of technological innovation that may serve as a unique intervention for youth experiencing psychological distress is the conversational agent, otherwise known as a chatbot. Further research is needed on the factors that may make mental health chatbots destined for adolescents more appealing and increase the likelihood that adolescents will use them.

Objective: The aim of this study was to assess adolescents' emotional reactions and likelihood of responding to questions that could be posed by a mental health chatbot. Understanding adolescent preferences and factors that could increase adolescents' likelihood of responding to chatbot questions could assist in future mental health chatbot design destined for youth.

Methods: We recruited 19 adolescents aged 14 to 17 years to participate in a study with a 2×2×3 within-subjects factorial design. Each participant was sequentially presented with 96 chatbot questions for a duration of 8 seconds per question. Following each presentation, participants were asked to indicate how likely they were to respond to the question, as well as their perceived affective reaction to the question. Demographic data were collected, and an informal debriefing was conducted with each participant.

Results: Participants were an average of 15.3 years old (SD 1.00) and mostly female (11/19, 58%). Logistic regressions showed that the presence of GIFs predicted perceived emotional valence (β=–.40, P<.001), such that questions without GIFs were associated with a negative perceived emotional valence. Question type predicted emotional valence, such that yes/no questions (β=–.23, P=.03) and open-ended questions (β=–.26, P=.01) were associated with a negative perceived emotional valence compared to multiple response choice questions. Question type also predicted the likelihood of response, such that yes/no questions were associated with a lower likelihood of response compared to multiple response choice questions (β=–.24, P=.03) and a higher likelihood of response compared to open-ended questions (β=.54, P<.001).

Conclusions: The findings of this study add to the rapidly growing field of teen-computer interaction and contribute to our understanding of adolescent user experience in their interactions with a mental health chatbot. The insights gained from this study may be of assistance to developers and designers of mental health chatbots.

(JMIR Hum Factors 2021;8(1):e24343) doi:10.2196/24343

KEYWORDS

chatbots; conversational agents; mental health; well-being; adolescents; user experience; user preferences

Introduction

Background

Psychological distress is defined as emotional suffering, characterized by symptoms of depression (ie, sadness, disinterest) and anxiety (ie, tension, agitation) [1]. Longitudinal studies tracking trajectories of psychological distress suggest that they increase across adolescence among both boys and girls [2-4]. Psychological distress has been associated in both meta-analytic and longitudinal studies with important health outcomes such as tobacco use [5,6], drug use [6], and alcohol use [7], with consequences that can extend into adulthood. As such, any interventions aimed at assisting adolescents who may be dealing with psychological distress are of high social importance to reduce their suffering and the consequences associated with distress. One type of technological innovation that may serve as a unique intervention for youth experiencing psychological distress is the conversational agent, otherwise known as a chatbot. Chatbots are "machine conversation systems [that] interact with human users via natural conversational language" [8]. Mental health chatbots are not only increasingly accessible and affordable but may also offer services to individuals who might not seek care due to stigma, elevated cost, or discomfort related to face-to-face therapy [9]. Mental health chatbots have been developed for use among clinical [10,11] and nonclinical [12-14] adult populations. Studies have shown that chatbot users experience improvement in psychological well-being, and tend to find the bots helpful and trustworthy [11,12]. Chatbots geared toward mental health are not only capable of identifying individuals who experience psychological distress but can also help reduce this distress [15]. Furthermore, these agents tend to be rated positively on measures of empathy and alliance [16].

Although chatbots may be deemed more suitable to adolescents, who are more familiar with smartphones [17], most studies on user-chatbot interactions have focused on adults. Among the few studies evaluating mental health chatbots geared toward helping youth, several indicate that these chatbots are effective in the detection and reduction of stress [18], anxiety, and depression [12,13,19]. One study showed that those who consistently interacted with the chatbot seemed to benefit from it [14], suggesting that increasing the likelihood of adherence such as by making these chatbots pleasant to use would be critical in their effectiveness. As such, the focus of this study was to evaluate the factors that increase adolescents' likelihood of responding to a mental health chatbot and which of its features they perceive more positively.

Related Research

Researchers have recognized the need for a better understanding of the behaviors and preferences of teens as they increasingly interact with technology [8] and, more specifically, chatbots [20,21]. A review of the literature on human-chatbot interaction highlighted the need for more user-centered research that aims to investigate how and why individuals choose to engage with a particular chatbot [22], as well as how they respond to it.

User experience includes perceptions and responses to the use of a product, as well as any emotions or preferences that occur during its use [23]. Personalizing a chatbot involves customizing its functionality, interface, and content to increase its relevance for an individual or group of individuals [24]. Adherence and engagement are essential to the success of mental health chatbots [25], as they are often associated with better outcomes [14,26]. A chatbot's characteristics may play an important role in determining whether users will regularly interact with it, as well as whether their interaction will be a pleasant one [27]. The propensity of users to voluntarily share information about themselves is especially crucial for favorable outcomes in their interactions with mental health chatbots. Although the acceptability and effectiveness of these chatbots have been explored [28], little attention has been paid to their characteristics (eg, language, personality) and how they impact user experience.

Two studies that have comprehensively investigated user experience with mental health chatbots described the design and development process of iHelpr, a chatbot that administers self-assessment scales and provides well-being information to adults [29]. The authors not only illustrated the design process but also reviewed the literature on user experience to outline a list of best practices for the design of chatbots in mental health care. Specifically, Cameron and colleagues [29] highlighted adapting the complexity of the chatbot's language to target users, and varying the content and conversation through the use of GIFs as some of the best practices for mental health chatbots. An evaluation of iHelpr's usability revealed that participants appreciated its friendly and upbeat personality, and also enjoyed the use of GIFs [29]. A chatbot's language and the use of graphics such as GIFs are only a few factors to consider when designing such technologies. Emojis, GIFs, and similar media can play a crucial role in determining the framework, sense, and direction of the conversation [25], and may also increase the social attractiveness and credibility of a chatbot [30].

Researchers are beginning to take interest in the effects of graphics on user interactions with mental health chatbots. Fadhil et al [25] showed that users preferred the use of emojis when the chatbot's questions pertained to their mental health. Duijst [31] reported that participants generally had positive reactions to emojis in a customer service chatbot, suggesting that adding emojis to the chatbot's dialogue may result in a more pleasant experience. However, some participants felt that combining emojis with a formal tone was strange and inconsistent. Indeed, younger users expressed a preference for a more casual tone, combined with just a few emojis. Adapting a chatbot's language to its context and users is therefore crucial to improving rapport and user engagement [32]. For instance, chatbots that are expected to be empathic, such as mental health chatbots, may elicit a more positive response from users by communicating in a friendly tone [33,34]. In the context of mental health, where an empathic chatbot would be rated more positively than a less empathic chatbot [35], the use of professional or polite language may be too neutral, possibly leading users to perceive the chatbot as uncaring or indifferent.

Study Objectives

This study was designed to answer the following question: What are adolescent users' reactions to questions posed by a mental

health chatbot? More specifically, the objective of this study was to evaluate adolescents' preferences (ie, emotional valence and likelihood of responding) regarding the formulation of questions that might be posed by a mental health chatbot. Preference is indicated by participants' affective reactions and the likelihood of response to the chatbot's statements. Given past research suggesting that individuals may prefer emojis and friendly tones in mental health chatbots, the questions presented to participants differed according to their tone (friendly or formal) and the presence of GIFs (present or absent). Questions also differed in type (yes/no, multiple response choice, or open-ended). These factors were chosen based both on past research on mental health chatbots [24,36] as well as the fact that they are easily malleable factors that may improve user experiences. We hypothesized that adolescents would show a preference for questions including GIFs and those with a friendly tone. As the chatbot's questions also differed according to their type (open-ended or closed), we sought to explore whether adolescents' preferences would vary in response to question type.

Methods

Recruitment

Given that the goal of this study was to assess user preferences for mental health chatbot communication among community adolescents, 19 adolescents aged 14 to 17 years were recruited from the general population via flyers and Facebook advertisements. Participants were informed about the study aims and voluntary participation, and each participant was given compensation of a total value of US $23.74.

Design and Procedure

This in-lab study was performed using a 2×2×3 within-subjects factorial design; the factors were presence of GIF (present vs absent), question tone (friendly vs formal), and question type (open-ended vs yes/no vs multiple response choice). Eight main questions were composed (Multimedia Appendix 1), each addressing a different theme centered around general well-being, including mood, stress management, and peer pressure. Each question was modified according to different combinations of each factor, yielding 12 variations for each of the 8 main statements and thus generating a total of 96 questions. The specific topic of each question was maintained across the different variations to control for the effect of theme on users' reactions. When comparing two levels of one experimental factor (eg, GIF present vs GIF absent), the same question was used for both conditions. The questions and GIFs were developed and pretested by four experts who were experienced in chatbot development. In addition, two adolescents were asked to provide feedback on the proposed questions prior to testing, commenting on readability and understanding of the questions. Sixteen GIFs were evaluated and the final eight (one per main question) were chosen by an expert panel. Sample questions are shown in Figures 1 and 2.

Once participants had read and signed the consent form, a research assistant explained the study rationale and gave participants brief verbal instructions. Participants were told to imagine that the questions presented to them were posed by a chatbot that aims to converse with users about their general well-being. Detailed instructions appeared on the computer screen at the start of the study. Participants were encouraged to take their time and to ask questions as needed to ensure they understood the task. All participants were also asked to complete a trial round before beginning the study. Data collection began once participants demonstrated a clear understanding of the task. Each of the 96 questions was presented sequentially on a computer screen for a duration of 8 seconds. Following each presentation, participants were automatically redirected to a short questionnaire presented via Qualtrics (USA) and asked to indicate their likelihood of responding to the question they had just read, as well as their perceived affective reaction to the question. The order of the chatbot questions was randomized for each participant. To prevent participant fatigue, a short 2-minute video was played after each set of 32 questions for a total of two video breaks. At the end of the study, we collected demographic data through another online questionnaire presented via Qualtrics. Informal debriefing was conducted at the end of data collection, and participant feedback was solicited and noted. Data collection lasted between 60 and 90 minutes per participant. An illustration of the study procedure is shown in Figure 3. This study received ethics approval from the Research Ethics Board of HEC Montreal.

Figure 1. Sample question (friendly tone, open-ended, with GIF).

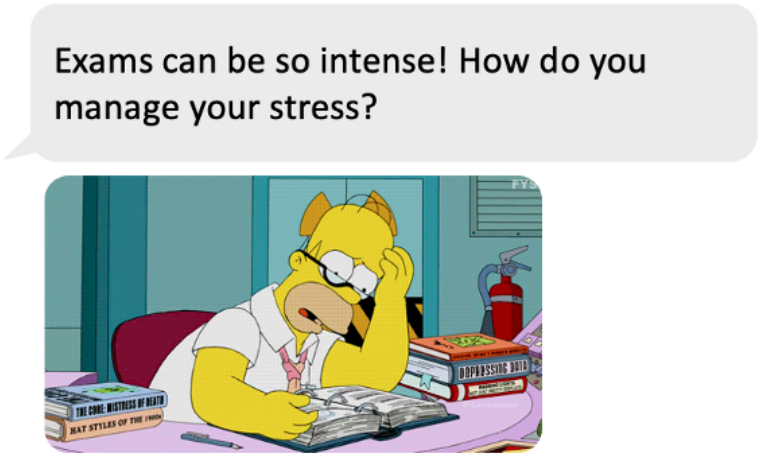


Figure 2. Sample question (professional tone, yes/no, without GIF).

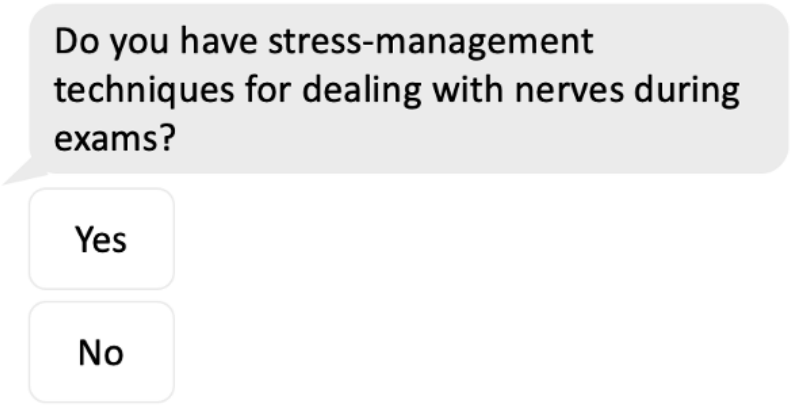
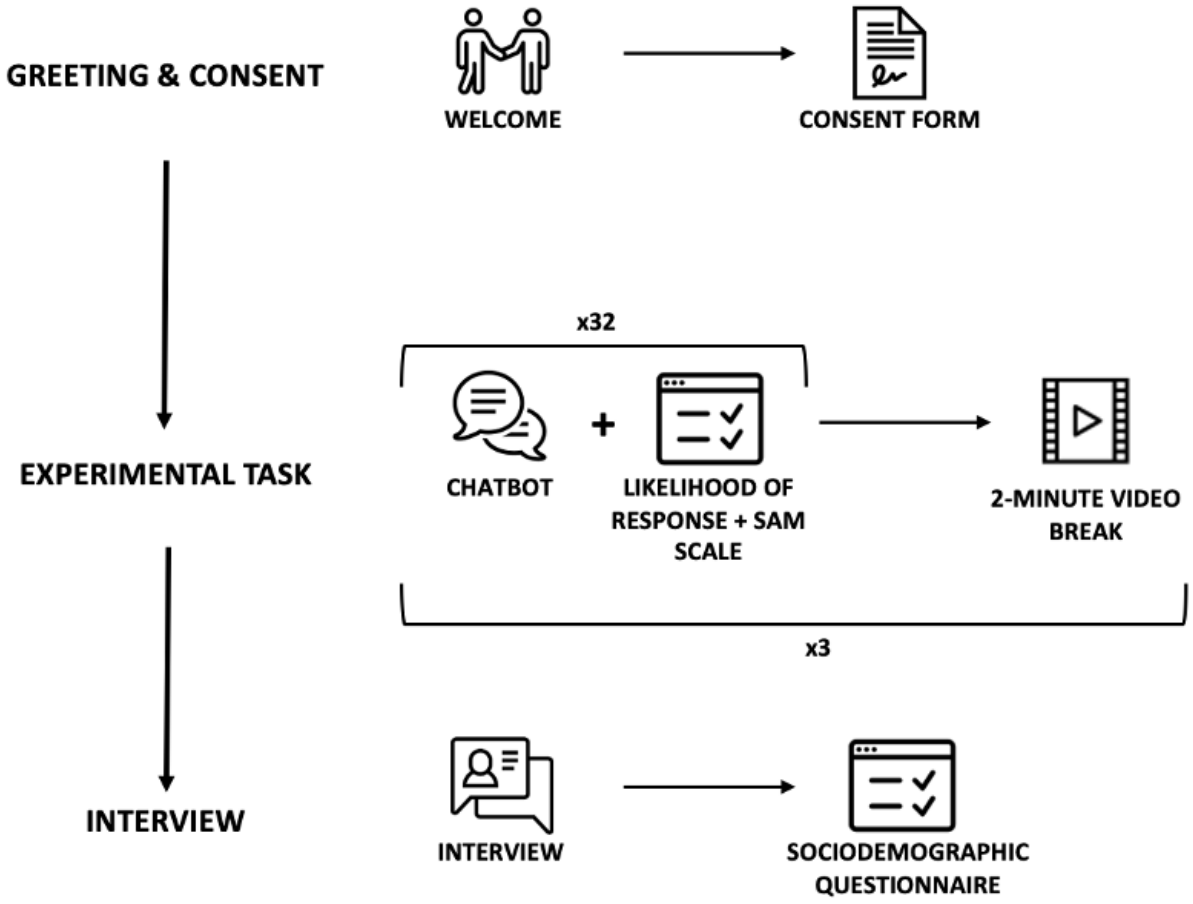


Figure 3. Study procedure. SAM: Self-Assessment Manikin.



Measures

Perceived Emotional Valence

The Self-Assessment Manikin scale is a 9-point nonverbal pictorial assessment tool used to measure the valence associated with one's affective reactions to stimuli [37]. Valence responses range from sad (1) to happy (9), with lower scores indicating negative valence and higher scores indicating positive valence.

Likelihood of Response

Participants' likelihood of responding to each question was measured with a 5-point Likert scale. Responses ranged from not at all likely (1) to very likely (5). See Multimedia Appendix 2 for the questionnaire used in this study.

Statistical Analysis

Due to the nonindependent nature of the observations (96 consecutive observations per participant), panel logistic regressions were performed to assess associations between the presence of a GIF, question type, question tone, and each

outcome (likelihood of response and perceived valence). Tests were performed while controlling for age, sex, presence of GIF, question type, and tone. The outcome variables were treated as ordinal variables. Regressions were carried out using SAS (version 9.4) and a posthoc power analysis was performed in R. Power analyses revealed that statistical power for the effects of a GIF, question type, and tone on perceived valence was 85%, which is satisfactory, with an odds ratio of 1.35 (β=.30). These analyses also revealed that statistical power for the effects of a GIF, question type, and tone on likelihood of response was 96% with an odds ratio of 1.49 (β=.40).

Results

Participant Demographics

Participants were an average of 15.3 years old (SD 1.00) and mostly female (11/19, 58%).

Perceived Emotional Valence

Question type significantly predicted perceived emotional valence, such that yes/no questions and open-ended questions were associated with a negative perceived emotional valence compared to multiple response choice questions. This suggests that participants had unpleasant affective reactions to yes/no and open-ended questions. Presence of a GIF also predicted perceived emotional valence, such that questions without GIFs were associated with a negative perceived emotional valence, suggesting that questions without GIFs were associated with negative affective reactions. Age group and sex (control variables) did not significantly predict emotional valence, and there was no statistically significant association between tone and perceived emotional valence (Table 1).

Table 1. Ordinal logistic regression for factors associated with perceived emotional valence (N=19).

Predictor comparison	β (SE)	*P* value
Presence vs absence (reference) of GIF	–.40 (.09)	<.001
Friendly vs professional (reference) tone	–.15 (.09)	.09
Question type		
Yes/No (reference) vs multiple response choice	–.23 (.10)	.03
Yes/No (reference) vs open-ended	.03 (.10)	.78
Multiple response choice (reference) vs open-ended	.26 (.11)	.01

Likelihood of Response

Question type significantly predicted the likelihood of response, such that yes/no questions were associated with a lower likelihood of response compared to multiple response choice questions and a higher likelihood of response compared to open-ended questions. Furthermore, multiple response choice questions were associated with a significantly higher likelihood of response compared to open-ended questions. Age group was a statistically significant predictor of likelihood of response (β=1.61, *P*=.02), whereas sex was not. Tone and presence of a GIF did not show statistically significant associations with likelihood of response (Table 2).

Table 2. Ordinal logistic regression for factors associated with likelihood of response (N=19).

Predictor comparison	β (SE)	*P* value
Presence vs absence (reference) of GIF	–.04 (.09)	.68
Friendly vs professional (reference) tone)	.06 (.09)	.48
Question type		
Yes/No (reference) vs multiple response choice	–.24 (.11)	.03
Yes/No (reference) vs open-ended	.54 (.11)	<.001
Multiple response choice (reference) vs open-ended	.78 (.11)	<.001

Discussion

Principal Findings

The objective of this study was to investigate adolescents' preferences regarding question formulation in the context of mental health chatbots. We hypothesized that adolescents would favor questions including GIFs as well as those with a friendly tone. We were also interested in observing whether adolescents preferred certain types of questions over others. Consistent with previous research [36], our results indicate that adolescents' self-reported affective reactions were significantly more positive in response to questions including GIFs compared to those without GIFs. With respect to question type, participants not only reported more positive affective reactions to multiple response choice questions compared to yes/no and open-ended questions but were also significantly more likely to respond to multiple response questions compared to other question types. The results show that the question features that elicited positive affective reactions did not necessarily lead to a high likelihood of response, and vice versa. For instance, although participants reacted positively to questions with GIFs, the inclusion of GIFs

had no statistically significant effect on the likelihood of response.

Participants' informal verbal feedback provides us with a more nuanced understanding of their experiences and preferences. As reflected in our findings, anecdotal evidence suggests that participants expressed a liking for the inclusion of GIFs in the chatbot's questions; although they found that GIFs added humor to certain questions, participants did not like all GIFs, and felt that some of these animated images were not relevant to the question with which they were paired. Thus, one possibility is that although participants reacted positively to the GIFs, such images may deter users from responding to certain questions if they are not deemed suitable to the chatbot's question.

Concerning question type, participants expressed an appreciation for closed questions. Although participants felt that open-ended questions allow them to better express themselves without feeling restricted by predetermined response choices, adolescents found closed questions "easier to respond to." Interestingly, despite the lack of statistically significant effects for question tone, participants shared positive reflections regarding the friendly tone. In fact, 10 participants specifically mentioned that they enjoyed the use of a friendly tone because it made the chatbot more "relatable" and "human-like," and 5 participants explicitly stated that they disliked questions with a formal tone. Nevertheless, several participants informally stated that when the chatbot's tone was overly friendly, it seemed as though the chatbot was "trying too hard." Furthermore, two participants preferred the formal tone to the friendly one; indeed, these participants felt that the formal tone was more appropriate to the types of questions being posed, whereas the friendly tone gave them the impression that they were not being taken seriously by the chatbot.

User Experience and Mental Health Chatbots

The findings of this study help us better understand user experience while interacting with a mental health chatbot. The participants' informal feedback highlights the variability within user preferences and reactions to the features of such chatbots. This variability has been observed in previous research. Yalcin and DiPaola [35] found that user interactions with M-Path, an empathic virtual agent, were not homogeneous. Furthermore, the authors observed that when participants showed more negative emotions, they rated the empathic agent more positively [35], thus illustrating an inconsistency in users' affective reactions to and ratings of the chatbot. Gaining a better understanding of the function of emotion within user experience is crucial to comprehending user-chatbot interaction, as emotion is closely tied to user acceptance and satisfaction [38] and influences motivation for consumptive behavior [39]. Furthermore, design guidelines for chatbots are generally heterogeneous and largely based on common knowledge rather than on empirical evidence [25]. More often than not, existing chatbots in various domains fail to meet consumer expectations, leading to user frustration and discontinued chatbot use [22,40]. Adolescents are indeed a heterogenous group in many respects and this heterogeneity can be illustrated by the different subcultures that exist among adolescents. Crutzen et al [41] suggest that "subculture-related differences should be taken into account while identifying user needs." An individual's personal characteristics also impact their preferences as well as their perceived value of and intention to use a given product. Therefore, to design successful products with specific target users, such as chatbots, developers should be guided by data from the user's point of view [42].

Limitations

Several limitations should be considered in the interpretation of these results. The results of this study reflect adolescents' reactions to potential questions posed by a mental health chatbot used in a voluntary fashion by community adolescents. Thus, these findings may not be generalizable to other chatbots such as customer service agents or mandatory use mental health chatbots. In addition, this study investigated only a few of the many features crucial to chatbot design. Moreover, the results may have been affected by decision fatigue, which can occur when sequential judgments need to be made within a certain time frame. Indeed, asking participants to make multiple ratings or to provide multiple responses in one session could impact subjective usability ratings [43]. However, we do not expect systematic biases in responding, given that the presentation of questions was random and video breaks were inserted into the study protocol. Breaks can be restorative and may "allow a return to original response levels" [44]. Lastly, although the questions and GIFs were pretested by experts, the pretest might have been more thorough if the questions had been pretested using quantitative methods (eg, rated by participants through a survey).

Conclusions and Future Research

In summary, this study evaluated adolescents' perceived emotional reactions and likelihood of response to questions posed by a mental health chatbot. These findings add to the rapidly growing field of teen-computer interaction and contribute to our understanding of adolescent user experience in their interactions with a mental health chatbot. A follow-up study should explore which characteristics of GIFs (eg, humor, relevance, size) might play a role in the identified effects, and how user reactions may vary based on different GIFs and based on the different questions posed (ie, the question themes). Future research might also observe users' back and forth conversations with a prototypical chatbot to investigate design elements that increase user satisfaction and that prolong interaction with the chatbot. The insights gained from this study may be of assistance to developers and designers of mental health chatbots geared toward adolescents. Employing an iterative design process is key to the optimization of mental health chatbots, and evaluating factors that increase user self-disclosure, engagement, and adherence are crucial to the success of these chatbots.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Undergraduate Student Research Award (USRA).

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of questions posed by the chatbot.
[PDF File (Adobe PDF File), 311 KB - humanfactors_v8i1e24343_app1.pdf]

Multimedia Appendix 2
Questionnaire presented following each chatbot question.
[PNG File , 323 KB - humanfactors_v8i1e24343_app2.png]

References

1. Mirowsky J, Ross CE. Measurement for a human science. J Health Soc Behav 2002 Jun;43(2):152-170. [Medline: 12096697]
2. Ge X, Conger RD, Elder GH. Pubertal transition, stressful life events, and the emergence of gender differences in adolescent depressive symptoms. Dev Psychol 2001 May;37(3):404-417. [doi: 10.1037//0012-1649.37.3.404] [Medline: 11370915]
3. Adkins DE, Wang V, Elder GH. Stress processes and trajectories of depressive symptoms in early life: Gendered development. Adv Life Course Res 2008 Jan;13:107-136. [doi: 10.1016/s1040-2608(08)00005-1]
4. Twenge JM, Nolen-Hoeksema S. Age, gender, race, socioeconomic status, and birth cohort differences on the children's depression inventory: a meta-analysis. J Abnorm Psychol 2002 Nov;111(4):578-588. [doi: 10.1037//0021-843x.111.4.578] [Medline: 12428771]
5. McKenzie M, Olsson C, Jorm A, Romaniuk H, Patton G. Association of adolescent symptoms of depression and anxiety with daily smoking and nicotine dependence in young adulthood: findings from a 10-year longitudinal study. Addiction 2010 Sep;105(9):1652-1659. [doi: 10.1111/j.1360-0443.2010.03002.x] [Medline: 20707783]
6. Hooshmand S, Willoughby T, Good M. Does the direction of effects in the association between depressive symptoms and health-risk behaviors differ by behavior? A longitudinal study across the high school years. J Adolesc Health 2012 Feb;50(2):140-147. [doi: 10.1016/j.jadohealth.2011.05.016] [Medline: 22265109]
7. Saraceno L, Heron J, Munafò M, Craddock N, van den Bree MBM. The relationship between childhood depressive symptoms and problem alcohol use in early adolescence: findings from a large longitudinal population-based study. Addiction 2012 Mar;107(3):567-577. [doi: 10.1111/j.1360-0443.2011.03662.x] [Medline: 21939461]
8. Shawar B, Atwell E. Chatbots: are they really useful? LDV Forum 2007;22(1):29-49.
9. Vaidyam A, Wisniewski H, Halamka J, Kashavan M, Torous J. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can J Psychiatry 2019 Jul;64(7):456-464 [FREE Full text] [doi: 10.1177/0706743719828977] [Medline: 30897957]
10. Tielman ML, Neerincx MA, Bidarra R, Kybartas B, Brinkman W. A therapy system for post-traumatic Stress disorder using a virtual agent and virtual storytelling to reconstruct traumatic memories. J Med Syst 2017 Aug;41(8):125 [FREE Full text] [doi: 10.1007/s10916-017-0771-y] [Medline: 28699083]
11. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul 01;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
12. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational Agent (Woebot): A Randomized Controlled Trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
13. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
14. Ly KH, Ly A, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
15. Philip P, Micoulaud-Franchi J, Sagaspe P, Sevin ED, Olive J, Bioulac S, et al. Virtual human as a new diagnostic tool, a proof of concept study in the field of major depressive disorders. Sci Rep 2017 Feb 16;7:42656. [doi: 10.1038/srep42656] [Medline: 28205601]
16. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. Patient Educ Couns 2005 Oct;59(1):21-30. [doi: 10.1016/j.pec.2004.09.008] [Medline: 16198215]

17. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
18. Huang J, Li Q, Xue Y, Cheng T, Xu S, Jia J, et al. TeenChat: A chatterbot system for sensing and releasing adolescents' stress. In: Yin X, Ho K, Zeng D, Aickelin U, Zhou R, Wang H, editors. International Conference on Health Information Science. HIS 2015. Lecture Notes in Computer Science, vol 9805. Cham: Springer; 2015.
19. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
20. Markopoulos P, Read J, Hoÿsniemi J, MacFarlane S. Child computer interaction: advances in methodological research. Cogn Tech Work 2007 Mar 27;10(2):79-81. [doi: 10.1007/s10111-007-0065-0]
21. Fitton D, Bell B, Little L, Horton M, Read J, Rouse M, et al. Working with teenagers in HCI research: a reflection on techniques used in the taking on the teenagers project. In: Little L, Fitton D, Bell B, Toth N, editors. Perspectives on HCI Research with Teenagers. Human-Computer Interaction Series. Cham: Springer; 2016:237-267.
22. Brandtzaeg PB, Følstad A. Chatbots: changing user needs and motivations. Interactions 2018 Aug 22;25(5):38-43. [doi: 10.1145/3236669]
23. Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems. Online Browsing Platform. 2019. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en [accessed 2020-07-20]
24. Fan H, Poole MS. What is personalization? Perspectives on the design and implementation of personalization in information systems. J Organ Comput Electron Commer 2006 Jan;16(3-4):179-202. [doi: 10.1080/10919392.2006.9681199]
25. Fadhil A, Schiavo G, Wang Y, Yilma B. The effect of emojis when interacting with conversational interface assisted health coaching system. : ACM; 2018 Presented at: 12th EAI international conference on pervasive computing technologies for healthcare; 2018; New York. [doi: 10.1145/3240925.3240965]
26. Mohr DC, Duffecy J, Ho J, Kwasny M, Cai X, Burns MN, et al. A randomized controlled trial evaluating a manualized TeleCoaching protocol for improving adherence to a web-based intervention for the treatment of depression. PLoS One 2013;8(8):e70086 [FREE Full text] [doi: 10.1371/journal.pone.0070086] [Medline: 23990896]
27. McTear M, Callejas Z, Griol D. The conversational interface: Talking to smart devices. Switzerland: Springer International Publishing; 2016.
28. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. Int J Med Inform 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
29. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Best practices for designing chatbots in mental healthcare – a case study on iHelpr. 2018 Presented at: 32nd International BCS Human Computer Interaction Conference (HCI); July 2018; Belfast URL: https://www.scienceopen.com/hosted-document?doi=10.14236/ewic/HCI2018.129 [doi: 10.14236/ewic/HCI2018.129]
30. Beattie A, Edwards AP, Edwards C. A bot and a smile: interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. Commun Stud 2020 Feb 16;71(3):409-427. [doi: 10.1080/10510974.2020.1725082]
31. Duijst D. Can we improve the user experience of chatbots with personalisation? University of Amsterdam MSc Thesis. ResearchGate. 2017. URL: https://www.researchgate.net/publication/318404775_Can_we_Improve_the_User_Experience_of_Chatbots_with_Personalisation [accessed 2020-07-20]
32. Liao Q, Davis M, Geyer W, Muller M, Shami N. What can you do? studying social-agent orientation and agent proactive interactions with an agent for employees. 2016 Jun Presented at: ACM Conference on Designing Interactive Systems; 2016; Brisbane, Australia. [doi: 10.1145/2901790.2901842]
33. Papadatou-Pastou M, Campbell-Thompson L, Barley E, Haddad M, Lafarge C, McKeown E, et al. Exploring the feasibility and acceptability of the contents, design, and functionalities of an online intervention promoting mental health, wellbeing, and study skills in Higher Education students. Int J Ment Health Syst 2019;13:51 [FREE Full text] [doi: 10.1186/s13033-019-0308-5] [Medline: 31367229]
34. Ghandeharioun A, McDuff D, Czerwinski M, Rowan K. Towards understanding emotional intelligence for behavior change chatbots. 2019 Jul 22 Presented at: 8th International Conference on Affective Computing and Intelligent Interaction (ACII); September 2019; Cambridge, UK. [doi: 10.1109/ACII.2019.8925433]
35. Yalçin Ö, DiPaola S. M-Path: A conversational system for the empathic virtual agent. 2019 Presented at: BICA: Biologically Inspired Cognitive Architectures; August 2019; Seattle, WA p. 597-607.
36. Cameron G, Cameron D, Megaw G, Bond R, Mulvenna M, O'Neill S, et al. Assessing the usability of a chatbot for mental health care. : Springer, Verlag; 2018 Presented at: 5th International Conference on Internet Science; October 2018; St. Petersburg p. 121-132. [doi: 10.1007/978-3-030-17705-8_11]
37. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
38. Erevelles S. The role of affect in marketing. J Bus Res 1998 Jul;42(3):199-215. [doi: 10.1016/s0148-2963(97)00118-5]

39. Hirschman EC, Holbrook MB. Hedonic consumption: emerging concepts, methods and propositions. J Market 1982;46(3):92. [doi: 10.2307/1251707]
40. Grudin J, Jacques R. Chatbots, humbots, and the quest for artificial general intelligence. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 2019; Glasgow, Scotland. [doi: 10.1145/3290605.3300439]
41. Crutzen R, de Nooijer J, Brouwer W, Oenema A, Brug J, de Vries NK. A conceptual framework for understanding and improving adolescents' exposure to Internet-delivered interventions. Health Promot Int 2009 Sep 10;24(3):277-284. [doi: 10.1093/heapro/dap018] [Medline: 19515716]
42. Borsci S, Kuljis J, Barnett J, Pecchia L. Beyond the user preferences: aligning the prototype design to the users' expectations. Hum Factors Man 2014 Dec 15;26(1):16-39. [doi: 10.1002/hfm.20611]
43. Robertson I, Kortum P. Extraneous factors in usability testing: evidence of decision fatigue during sequential usability judgments. 2018 Sep 27 Presented at: Human Factors and Ergonomics Society Annual Meeting; September 1, 2018; Philadelphia, PA p. 1409-1413. [doi: 10.1177/1541931218621321]
44. Robertson I, Kortum P. The effect of cognitive fatigue on subjective usability scores. 2017 Oct 20 Presented at: Human Factors and Ergonomics Society Annual Meeting; September 1, 2017; Austin, TX p. 1461-1465. [doi: 10.1177/1541931213601850]

Edited by A Kushniruk; submitted 16.09.20; peer-reviewed by S Morana, A Brendel, R Krukowski, M Slinowsky; comments to author 01.11.20; revised version received 27.12.20; accepted 17.01.21; published 18.03.21.

Please cite as:
Mariamo A, Temcheff CE, Léger PM, Senecal S, Lau MA
Emotional Reactions and Likelihood of Response to Questions Designed for a Mental Health Chatbot Among Adolescents: Experimental Study
JMIR Hum Factors 2021;8(1):e24343
URL: https://humanfactors.jmir.org/2021/1/e24343
doi:10.2196/24343
PMID:33734089

Original Paper

Prevalence of Misinformation and Factchecks on the COVID-19 Pandemic in 35 Countries: Observational Infodemiology Study

Meeyoung Cha[1,2], DPhil; Chiyoung Cha[3], DPhil; Karandeep Singh[2], DPhil; Gabriel Lima[1,2], BSc; Yong-Yeol Ahn[4,5,6], DPhil; Juhi Kulshrestha[7], DPhil; Onur Varol[8], DPhil

[1]School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[2]Data Science Group, Institute for Basic Science, Daejeon, Republic of Korea

[3]College of Nursing, Ewha Womans University, Seoul, Republic of Korea

[4]Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States

[5]Indiana University Network Science Institute, Indiana University, Bloomington, IN, United States

[6]Connection Science, Massachusetts Institute of Technology, Cambridge, MA, United States

[7]GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany

[8]Computer Science Department, Sabanci University, Istanbul, Turkey

Corresponding Author:
Meeyoung Cha, DPhil
School of Computing
Korea Advanced Institute of Science and Technology
291 Daehak-ro Yuseong-gu
Daejeon
Republic of Korea
Phone: 82 8789300
Email: mcha@ibs.re.kr

Abstract

Background: The COVID-19 pandemic has been accompanied by an infodemic, in which a plethora of false information has been rapidly disseminated online, leading to serious harm worldwide.

Objective: This study aims to analyze the prevalence of common misinformation related to the COVID-19 pandemic.

Methods: We conducted an online survey via social media platforms and a survey company to determine whether respondents have been exposed to a broad set of false claims and fact-checked information on the disease.

Results: We obtained more than 41,000 responses from 1257 participants in 85 countries, but for our analysis, we only included responses from 35 countries that had at least 15 respondents. We identified a strong negative correlation between a country's Gross Domestic Product per-capita and the prevalence of misinformation, with poorer countries having a higher prevalence of misinformation (Spearman ρ=–0.72; P<.001). We also found that fact checks spread to a lesser degree than their respective false claims, following a sublinear trend (β=.64).

Conclusions: Our results imply that the potential harm of misinformation could be more substantial for low-income countries than high-income countries. Countries with poor infrastructures might have to combat not only the spreading pandemic but also the COVID-19 infodemic, which can derail efforts in saving lives.

(JMIR Hum Factors 2021;8(1):e23279) doi:10.2196/23279

KEYWORDS

COVID-19; coronavirus; infodemic; infodemiology; misinformation; vulnerability; LMIC countries

Introduction

COVID-19, caused by SARS-CoV-2 [1], has spread worldwide, becoming a global pandemic. Most preventive measures against the disease comprise individual behaviors, as therapeutics are under development and yet to be approved by national health agencies [2]. Since such measures require individuals to change their behaviors according to validated information about the disease, effective communication of accurate information to the

public is critical for minimizing the pandemic's impact. However, as observed previously in the context of the anti-vaccination movement [3], communicating accurate health-related information can be challenging. Moreover, social media can rapidly disseminate a piece of misinformation about the disease to millions of people. The amplifying nature of online platforms can threaten public health by creating an *infodemic* [4].

The COVID-19 infodemic has shown to be exceptionally harmful in the context of both individual and public health. For instance, misinformation has motivated people to attack and abuse health care workers [5]. There have been reports of eggs being thrown at nurses in Mexico [6] and Indian doctors being evicted from their houses under the belief that they were vectors of the disease [7]. An even worse event has occurred in Iran, where over a hundred people died and thousands became severely ill due to methanol poisoning [8]. The infodemic has also had negative consequences on the psychosocial health of various layers of society. Widespread misinformation has affected the global population by increasing levels of depression, anxiety, and posttraumatic stress disorder [9].

The negative health impact of the infodemic could also widen the already extensive world health literacy gap. People with low health literacy are more vulnerable to an infodemic, as they tend to have a limited ability to seek, comprehend, and evaluate health information from social media [10]. Research has found a strong relationship between health literacy and adverse health outcomes from infectious diseases [11]. Therefore, assessing the level of eHealth literacy on COVID-19 and understanding how the COVID-19 infodemic has spread worldwide are crucial efforts in the current pandemic [12].

Many of the infodemic's claims have been locally fact-checked. However, debunking information does not spread effectively and rapidly enough through the global population, leading to misinformation and causing further harm in other parts of the world [13]. Prior research has shown that false news spreads faster than its factual counterparts [14]. Hence, misinformation could gain a strong foothold over their trustworthy counterparts, and the current COVID-19 infodemic might prove to be especially harmful, as it tackles health-related behaviors that could lead to life and death consequences.

To prevent the harm caused by misinformation, we have launched "Facts Before Rumors." This has been a pre-emptive public communication campaign to combat COVID-19 misinformation by spreading fact-checked information from countries that have seen false information earlier to regions that have not necessarily seen the same piece of information yet. Our campaign's distinguishing feature is that the current project is proactively propagating validated responses to claims seen in other countries and regions, thereby pre-emptively suppressing the dissemination of false health information. Alongside our campaign, we have also conducted a survey-based study to quantify the infodemic's reach worldwide. Specifically, we investigated the public exposure to false claims and fact-checked information across different world regions. We present our findings in the following sections.

Methods

After identifying more than 200 claims about COVID-19 that had been fact-checked in China, duplicated claims, rumors not related to health (eg, political conspiracies), and claims addressing local topics were removed. Two Chinese-speaking researchers were involved in this process. In total, 11 pieces of misinformation addressing health-related behaviors were chosen:

1. Hot: The virus will only spread in cold, dry weather and does not survive in hot, humid weather.
2. Sauna: Hot baths or saunas can reduce the chances of getting infected with COVID-19.
3. Drink: Drinking water or tea frequently will cure a COVID-19 infection.
4. Mask: Microwave, steam, blow-dry, or spray alcohol to clean used face masks.
5. Garlic: Garlic, ginger, onion, sesame oil, probiotics, herbal remedies, or aromatherapy can prevent the infection.
6. Dryer: Hot air dryers can kill the virus.
7. Salt: Gargling with salt water, vinegar, or saline nose rinse can eliminate the virus.
8. Age: Only certain age groups, races, and ethnicities are vulnerable to the virus.
9. Test: You can test yourself for COVID-19 by holding your breath for 10 seconds.
10. Eggs: Eating eggs every day can cure the virus.
11. Bleach: Spraying alcohol or chlorine over your body will kill the virus.

We conducted a large-scale online survey via Pollfish, a survey company, and personal social media channels. The respondents were recruited via convenience sampling, given the large-scale nature of the study. Pollfish conducts surveys by randomizing its delivery to the targeted populations via mobile apps. Respondents were compensated with nonmonetary incentives such as extra lives in a game or access to premium content. As per the documentation, Pollfish has partnerships with over 120,000 app providers and is present in over 160 countries worldwide. We obtained more than 41,000 responses from 1257 unique individuals residing in 85 countries between early April and mid-May 2020.

In our study, we asked whether participants have seen the chosen claims, whether they believed that exposing fact-checked information of those claims would be beneficial to their community, and whether these claims have been either confirmed or denied by official sources. The respondents also reported their perceived financial and health status, alongside various demographic questions, such as sex and age. For analysis, we only included 35 countries that had at least 15 respondents to eliminate noisy and biased observations for those countries.

Participant recruitment relied on the survey platform's methodology, and the only demographic control added was age (ie, older than 18 years). To the survey question "How would you rate your financial status?" participants on average reported a similar level of perceived financial status near the response category "fair" among "very poor," "poor," "fair," "good," and "excellent," (mean 0.268, median 0, when converted to a 5-point

bipolar scale). Hence, we consider that respondents from distinct countries belong to similar economic classes. We report the demographic distribution of survey participants in Table 1.

Table 1. Demographic attributes of survey participants (N=1257).

Demographic attributes	Participants, n (%)
Gender	
Female	499 (39.70)
Male	750 (59.67)
Other	8 (0.63)
Age groups (years)	
18-24	399 (31.74)
25-34	422 (33.57)
35-44	282 (22.43)
44-54	97 (7.71)
55-64	48 (3.81)
≥65	9 (0.70)
Education	
High school	521 (41.45)
University or college	409 (32.53)
Graduate school or more	327 (26.01)
Health status	
Very poor	14 (1.11)
Poor	42 (3.34)
Fair	205 (16.30)
Good	616 (49.01)
Very good	380 (30.23)
Financial status	
Very poor	47 (3.74)
Poor	169 (13.44)
Fair	531 (42.24)
Good	420 (33.41)
Excellent	90 (7.16)

Results

Although some false claims were geographically confined, our survey revealed that many false claims recurred across different countries and languages, highlighting the far-reaching power of the infodemic.

For instance, the claim stating that eating eggs every day could cure the disease has primarily spread across Asia. In contrast, the claim stating that SARS-CoV-2 would only spread under cold and dry weather was seen (in its entirety or partly) by more than 82% of our total respondents across all continents.

When comparing the exposure to false claims across countries, we found that countries with lower gross domestic product per capita, which are likely much more vulnerable to the disease itself, tend to exhibit higher rates of exposure to false claims (see Figure 1; Spearman ρ=–0.72; P<.001). Our observation indicates that these false claims could particularly hit countries or groups of people with limited access to information even harder, compounding the finding that poorer countries are more vulnerable to communicable diseases [15].

Figure 1. Exposure rate of the selected 11 claims across different countries. Results are shown for countries with at least 15 survey participants. The x-axis indicates the log of GDP per capita of different countries. The countries in the increasing order of GDP per capita are Ethiopia, Nepal, Pakistan, Cambodia, Bangladesh, Kenya, India, Nicaragua, Nigeria, Egypt, Philippines, Indonesia, Sri Lanka, Azerbaijan, South Africa, Thailand, Cuba, Brazil, Turkey, Russia, Argentina, Romania, Chile, Venezuela, Estonia, Bahrain, South Korea, Italy, the United Kingdom, the United Arab Emirates, Canada, Germany, Finland, Sweden, and the United States. The y-axis indicates the average percentage of claims that respondents had seen (ie, mean exposure to claims in a country). GDP: gross domestic product.

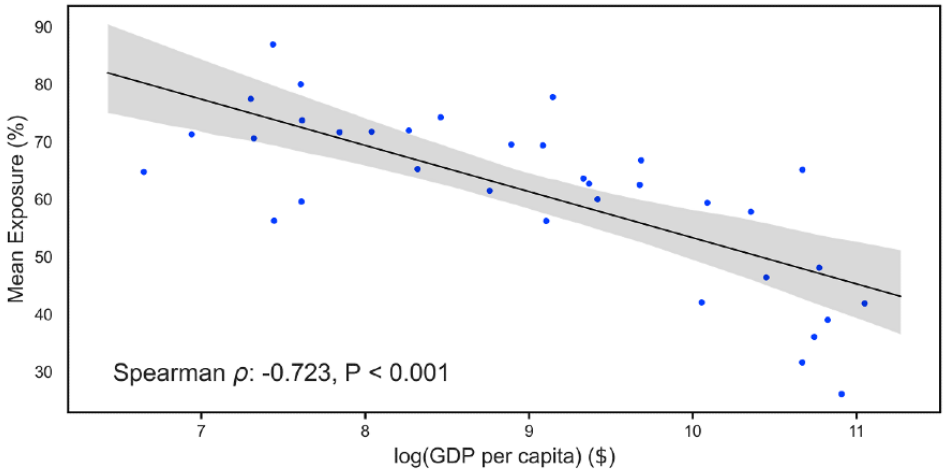
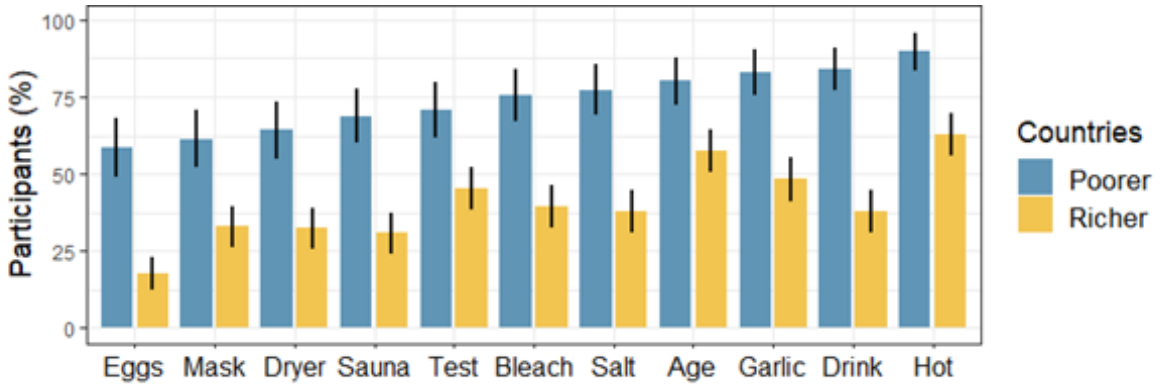


Figure 2 depicts the average exposure rates for the eleven claims for the top and bottom five countries ranked by gross domestic product per capita. It exemplifies how different rumors have distinct spread patterns worldwide. Unlike the actual disease, which first invaded richer countries, the infodemic that follows it is attacking, via the well-connected internet communities, vulnerable countries the most.

Figure 2. Exposure rate of the selected 11 claims in the poorest and wealthiest countries. Rates are shown for the bottom (Ethiopia, Nepal, Pakistan, Cambodia, Bangladesh) and top five (the United States, Sweden, Finland, Germany, Canada) countries in terms of gross domestic product per capita. The bars depict the mean percentage of respondents who have at least partially seen each claim and standard error bars. Claims on the x-axis are sorted by mean claim exposure rate in the poorest countries.



Our results also indicated that those countries with a higher incidence of false claims have also had a substantial amount of fact checks debunking this information. However, this relationship is sublinear; for an increase of 1% in citizens seeing our selected claims, marginally more than half (β=.64) of them would have also been presented with debunking information. Therefore, countries most affected by the COVID-19 misinformation rapidly encounter false information that is not fact-checked by official sources at the same rate. This is a concern as people are less active in seeking personal health strategies [16].

Furthermore, people who had been previously exposed to COVID-19 claims perceived a more significant benefit in sharing fact-checked information of claims (Pearson r=0.44; P<.001). This means that campaigns such as ours would be viewed as valuable, particularly in countries currently experiencing misinformation at a higher degree.

Discussion

Our results highlight that low-income countries may be at a higher risk of exposure to misinformation and be disadvantaged by the COVID-19 infodemic during the global pandemic. Fact-checked information does not propagate at the same speed as false information, and therefore, countries most affected by the infodemic should also have a higher incidence of unchecked information. Our results warrant a pre-emptive strategy for busting misinformation and indicate a higher demand for localized fact checks in these countries and a public belief, especially in low-income countries, that fact-checking campaigns can benefit their local community.

The analysis presented in this study has some limitations. Although we conducted a large-scale survey to quantify the spread of misinformation in different countries, our sample was not necessarily representative of the target countries' population. Additionally, our survey covered respondents from 35 countries, and our analysis did not consider other parts of the world. Moreover, we have included 11 health-related false claims that circulated online in China during the pandemic's infancy. Future work could address a broader range of rumors, such as from political topics, as we have chosen to not tackle them in this study. We have also not obtained information about how respondents were exposed to claims, such as through social media platforms or traditional media.

In spite of the aforementioned limitations, the current analysis results have yielded interesting and useful insights into the spread of misinformation across different countries of the world. As future work, we plan to propagate our campaign to a broader audience in more countries to suppress the infodemic proactively and extend our results from this and upcoming studies to more representative samples and other demographic variables.

Acknowledgments

GL, KS, and MC are supported by the Institute for Basic Science in South Korea (IBS-R029-C2). MC is also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF-2017R1E1A1A01076400). YYA is supported by the Defense Advanced Research Projects Agency (DARPA), contract W911NF-17-C-0094.

Conflicts of Interest

None declared.

References

1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet 2020 Feb 15;395(10223):470-473 [FREE Full text] [doi: 10.1016/S0140-6736(20)30185-9] [Medline: 31986257]
2. Chinazzi M, Davis JT, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science 2020 Apr 24;368(6489):395-400 [FREE Full text] [doi: 10.1126/science.aba9757] [Medline: 32144116]
3. Dubé E, Vivion M, MacDonald NE. Vaccine hesitancy, vaccine refusal and the anti-vaccine movement: influence, impact and implications. Expert Rev Vaccines 2015 Jan;14(1):99-117. [doi: 10.1586/14760584.2015.964212] [Medline: 25373435]
4. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for Managing the COVID-19 Infodemic: Methods and Results of an Online, Crowdsourced WHO Technical Consultation. J Med Internet Res 2020 Jun 26;22(6):e19659 [FREE Full text] [doi: 10.2196/19659] [Medline: 32558655]
5. McKay D, Heisler M, Mishori R, Catton H, Kloiber O. Attacks against health-care personnel must stop, especially as the world fights COVID-19. Lancet 2020 Jun 06;395(10239):1743-1745 [FREE Full text] [doi: 10.1016/S0140-6736(20)31191-0] [Medline: 32445692]
6. Semple K. 'Afraid to Be a Nurse': health workers under attack. The New York Times. 2020. URL: https://www.nytimes.com/2020/04/27/world/americas/coronavirus-health-workers-attacked.html [accessed 2020-11-12]
7. Withnall A. Coronavirus: why India has had to pass new law against attacks on healthcare workers. The Independent. 2020. URL: https://www.independent.co.uk/news/world/asia/coronavirus-india-doctor-nurse-attack-law-modi-health-care-workers-a9480801.html [accessed 2020-11-12]
8. Haghdoost Y. Alcohol poisoning kills 100 Iranians seeking virus protection. Bloomberg. 2020. URL: https://www.bloomberg.com/news/articles/2020-03-18/alcohol-poisoning-kills-100-iranians-seeking-virus-protection [accessed 2020-11-12]
9. Dubey S, Biswas P, Ghosh R, Chatterjee S, Dubey MJ, Chatterjee S, et al. Psychosocial impact of COVID-19. Diabetes Metab Syndr 2020;14(5):779-788 [FREE Full text] [doi: 10.1016/j.dsx.2020.05.035] [Medline: 32526627]
10. 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 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
11. Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461 [FREE Full text] [doi: 10.1177/1474515120941307] [Medline: 32667217]
12. Chong YY, Cheng HY, Chan HYL, Chien WT, Wong SYS. COVID-19 pandemic, infodemic and the role of eHealth literacy. Int J Nurs Stud 2020 Aug;108:103644 [FREE Full text] [doi: 10.1016/j.ijnurstu.2020.103644] [Medline: 32447127]
13. Donovan J. Social-media companies must flatten the curve of misinformation. Nature 2020 Apr 14. [doi: 10.1038/d41586-020-01107-z] [Medline: 32291410]
14. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Mar 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
15. Kankeu HT, Saksena P, Xu K, Evans DB. The financial burden from non-communicable diseases in low- and middle-income countries: a literature review. Health Res Policy Syst 2013 Aug 16;11:31 [FREE Full text] [doi: 10.1186/1478-4505-11-31] [Medline: 23947294]

16. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A 2020 May 26;117(21):11220-11222 [FREE Full text] [doi: 10.1073/pnas.2005335117] [Medline: 32366658]

Edited by A Kushniruk; submitted 06.08.20; peer-reviewed by Y Senathirajah, D Chrimes, H Monkman; comments to author 02.09.20; revised version received 28.10.20; accepted 23.12.20; published 13.02.21.

Please cite as:
Cha M, Cha C, Singh K, Lima G, Ahn YY, Kulshrestha J, Varol O
Prevalence of Misinformation and Factchecks on the COVID-19 Pandemic in 35 Countries: Observational Infodemiology Study
JMIR Hum Factors 2021;8(1):e23279
URL: https://humanfactors.jmir.org/2021/1/e23279
doi:10.2196/23279
PMID:33395395

Original Paper

Collaboration Structures in COVID-19 Critical Care: Retrospective Network Analysis Study

Chao Yan[1], MSc; Xinmeng Zhang[1], BSc; Cheng Gao[2], PhD; Erin Wilfong[3], MD, PhD; Jonathan Casey[3], MSci, MD; Daniel France[4], PhD, MPH; Yang Gong[5], MD, PhD; Mayur Patel[6,7,8,9], MPH, MD; Bradley Malin[1,2,10], PhD; You Chen[1,2], PhD

[1]Department of Electrical Engineering & Computer Science, Vanderbilt University, Nashville, TN, United States
[2]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States
[3]Division of Allergy, Pulmonary and Critical Care Medicine, Vanderbilt University Medical Center, Nashville, TN, United States
[4]Department of Anesthesiology, Vanderbilt University Medical Center, Nashville, TN, United States
[5]School of Biomedical Informatics, University of Texas Health Science Center at Houston, Houston, TX, United States
[6]Department of Hearing & Speech Sciences, Vanderbilt University Medical Center, Nashville, TN, United States
[7]Geriatric Research and Education Clinical Center, Veteran Affairs Tennessee Valley Healthcare System, Nashville, TN, United States
[8]Department of Surgery, Vanderbilt University Medical Center, Nashville, TN, United States
[9]Department of Neurosurgery, Vanderbilt University Medical Center, Nashville, TN, United States
[10]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:
You Chen, PhD
Department of Biomedical Informatics
Vanderbilt University Medical Center
2525 West End Ave
Nashville, TN, 37203
United States
Phone: 1 615 343 1939
Email: you.chen@vanderbilt.edu

Abstract

Background: Few intensive care unit (ICU) staffing studies have examined the collaboration structures of health care workers (HCWs). Knowledge about how HCWs are connected to the care of critically ill patients with COVID-19 is important for characterizing the relationships among team structures, care quality, and patient safety.

Objective: We aimed to discover differences in the teamwork structures of COVID-19 critical care by comparing HCW collaborations in the management of critically ill patients with and without COVID-19.

Methods: In this retrospective study, we used network analysis methods to analyze the electronic health records (EHRs) of 76 critically ill patients (with COVID-19: n=38; without COVID-19: n=38) who were admitted to a large academic medical center, and to learn about HCW collaboration. We used the EHRs of adult patients who were admitted to the COVID-19 ICU at the Vanderbilt University Medical Center (Nashville, Tennessee, United States) between March 17, 2020, and May 31, 2020. We matched each patient according to age, gender, and their length of stay. Patients without COVID-19 were admitted to the medical ICU between December 1, 2019, and February 29, 2020. We used two sociometrics—eigencentrality and betweenness—to quantify HCWs' statuses in networks. Eigencentrality characterizes the degree to which an HCW is a core person in collaboration structures. Betweenness centrality refers to whether an HCW lies on the path of other HCWs who are not directly connected. This sociometric was used to characterize HCWs' broad skill sets. We measured patient staffing intensity in terms of the number of HCWs who interacted with patients' EHRs. We assessed the statistical differences in the core and betweenness statuses of HCWs and the patient staffing intensities of COVID-19 and non–COVID-19 critical care, by using Mann-Whitney U tests and reporting 95% CIs.

Results: HCWs in COVID-19 critical care were more likely to frequently work with each other (eigencentrality: median 0.096) than those in non–COVID-19 critical care (eigencentrality: median 0.057; P<.001). Internal medicine physicians in COVID-19 critical care had higher core statuses than those in non–COVID-19 critical care (P=.001). Nurse practitioners in COVID-19 care had higher betweenness statuses than those in non–COVID-19 care (P<.001). Compared to HCWs in non–COVID-19 settings,

the EHRs of critically ill patients with COVID-19 were used by a larger number of internal medicine nurse practitioners (P<.001), cardiovascular nurses (P<.001), and surgical ICU nurses (P=.002) and a smaller number of resident physicians (P<.001).

Conclusions: Network analysis methodologies and data on EHR use provide a novel method for learning about differences in collaboration structures between COVID-19 and non–COVID-19 critical care. Health care organizations can use this information to learn about the novel changes that the COVID-19 pandemic has imposed on collaboration structures in urgent care.

(JMIR Hum Factors 2021;8(1):e25724) doi:10.2196/25724

KEYWORDS

COVID-19; intensive care unit; collaboration structure; critically ill patient; health care worker; network analysis; electronic health record; collaboration; critical care; relationship; safety; teamwork

Introduction

The COVID-19–Associated Hospitalization Surveillance Network has reported that the overall cumulative COVID-19 hospitalization rate in the United States is 199.8 people per 100,000 people for the week that ended October 24, 2020 [1]. Additionally, between 5% and 12.2% of patients aged <60 years and between 27.4% and 70.9% of patients aged ≥60 years have required intensive care due to deteriorating respiratory conditions [2-4]. Health care organizations (HCOs) have been exploring various approaches (eg, the creation of COVID-19 intensive care units [ICUs] and the extension of existing ICUs) for satisfying the increasing medical needs of critically ill patients with COVID-19 [5-7]. Various staffing strategies and protocols for the care of critically ill patients with COVID-19 have been developed [8-10]. These strategies belong to three ICU staffing model categories—open, closed, and hybrid. In an open model, many different medical staff members manage patients in ICUs. In contrast, the closed model limits the staffing system to ICU-certified physicians (eg, intensivists). The hybrid model draws upon the aspects of the open and closed models by staffing ICUs with an attending physician and a team so that they can work in tandem with primary physicians.

ICU staffing (eg, the assignment of patients to a set of health care workers [HCWs]) can impact care quality and patient safety [11-13]. As such, HCOs need to be mindful of how they assess collaborations among HCWs to properly care for critically ill patients with COVID-19. However, COVID-19 ICU staffing strategies are designed at a very high level (eg, team scheduling). Therefore, they neglect the cross-disciplinary connections among HCWs. Gaining knowledge on how HCWs connect and collaborate can improve teamwork, which in turn may improve care quality and patient safety [14].

Few studies have investigated the collaboration structures of COVID-19 critical care [15], but to the best of our knowledge, none have examined the collaborations among HCWs. As such, there is a limited amount of explicitly documented evidence about cross-disciplinary (eg, internal medicine physicians, respiratory therapists, and cardiovascular nurses) collaboration in COVID-19 critical care. HCOs need this information to manage teamwork and improve care quality and patient safety during the pandemic. In this study, we used network analysis methods to learn about the collaboration structures of COVID-19 critical care. We specifically investigated how HCWs are connected in the context of providing care to critically ill patients with COVID-19. One of the challenges in modeling the connections among HCWs in the ICU is their complexity (eg, cooperation among multidisciplinary HCWs). In this study, we learned about the collaborations among multidisciplinary HCWs by analyzing electronic health record (EHR) systems. EHR systems provide an environment that aids with teamwork (eg, the exchange of health information among HCWs). This can help HCOs with offering more accurate, detailed, and timely information, which would result in the delivery of higher quality care [16,17]. As EHR adoption has spread, the proportion of HCW activities (eg, the review of notes, requests for x-rays, and the management of medication) that involve EHRs has increased [18,19]. Thus, interactions with EHRs provide an opportunity for studying the collaborations among HCWs [20-24].

We conducted a secondary analysis on EHR use to learn about the collaborations among HCWs. We created networks by identifying connections among HCWs who conducted activities with the EHRs of the same patients on the same day.

We used two sociometrics—eigencentrality and betweenness centrality—to measure the core and betweenness status of an HCW in the collaboration network. Eigencentrality characterizes the degree to which an HCW is a core person in collaboration structures. Betweenness centrality refers to whether an HCW lies on the path of other HCWs who are not directly connected. An HCW who has a broad skill set and cares for a wide spectrum of patients could frequently be in a high-betweenness position.

We analyzed data on EHR system use in the Vanderbilt University Medical Center, which is a large academic medical center in Nashville, Tennessee that created its COVID-19 unit in the middle of March 2020. The high density of clinical ICU data and large volume of EHR activities for each ICU patient episode allow for the investigation of HCW collaboration in the management of critically ill patients before and during the COVID-19 pandemic.

We learned about the collaboration structures in COVID-19 critical care by comparing structures that were associated with the management of critically ill patients with and without COVID-19.

Methods

Data Set

We screened for adult patients who were admitted to the COVID-19 ICU between March 17, 2020, and May 31, 2020. We matched each patient with COVID-19 with an adult patient

without COVID-19 who was admitted to the medical ICU (MICU) between December 1, 2019, and February 29, 2020, via propensity score matching.

The propensity score was based on age, gender, and patients' length of stay. The distribution of the COVID-19 and non–COVID-19 groups' propensity scores is depicted in Multimedia Appendix 1. The Pearson correlation coefficient between the two distributions was 0.93; the associated *P* value was <.001. This proved that the variance in the confounding factors between the two patient groups was very small. We focused on patients who were alive at discharge because their hospital stays were relatively complete. This process yielded a sample of 76 critically ill, adult patients—38 with COVID-19 and 38 without COVID-19. In total, 3 patients with COVID-19 required multiple ICU stays. For this study, we randomly selected one stay for each of these patients. Table 1 provides a summary of the demographic characteristics, comorbidities, and outcomes of the investigated patients with and without COVID-19.

Table 1. Characteristics of the critically ill patients in this study.

Characteristics	Patients with COVID-19[a]	Patients without COVID-19[b]
Patients, n	38	38
Demographic characteristics		
Age (years), median (IQR; SD)	54 (47-66; 14)	54 (49-64; 12)
Sex, n (%)		
Female	15 (39)	15 (39)
Male	23 (61)	23 (61)
Race, n (%)		
White	22 (58)	32 (84)
African American	6 (16)	5 (13)
Asian	4 (11)	0 (0)
Other	6 (16)	1 (3)
Outcomes		
Length of stay (days), median (IQR; SD)	13.5 (6.50-18.75; 10)	13.5 (7.50-19.00; 9)
Hospital discharge disposition, n (%)		
Home	29 (76)	24 (63)
Other	9 (24)	14 (37)
Comorbidities, n (%)		
Hypertension	19 (50)	33 (87)
Cardiovascular disease	14 (37)	23 (61)
Renal disease	19 (50)	22 (58)
Diabetes	10 (26)	16 (42)
Chronic metabolic disease	14 (37)	18 (47)
Chronic lung disease	9 (24)	17 (45)

[a]Patients with COVID-19 were admitted to the intensive care unit between March 17, 2020, and May 31, 2020.

[b]Patients without COVID-19 were admitted to the medical intensive care unit between December 1, 2019, and February 29, 2020.

The study population had several notable aspects. First, we noticed that there was a disproportionate number of males. Second, while there were more self-reported White patients than patients of other races, the number of White patients in the COVID-19 group was substantially smaller than that in the non–COVID-19 group. Third, patients without COVID-19 had a high incidence of comorbidities; specifically, patients without COVID-19 exhibited the six comorbidities that are common in patients with COVID-19 (ie, those reported by the COVID-19–Associated Hospitalization Surveillance Network) [1]. Fourth, the majority of patients from the two groups (with COVID-19: 29/38, 76%; without COVID-19: 24/38, 63%) were discharged home.

Study Design

The analysis consisted of two primary components. First, we used network analysis methods to learn about the HCW networks that were involved in the management of critically ill patients. Second, we statistically compared and contrasted the network structures in COVID-19 and non–COVID-19 settings.

Modeling HCW Networks

We analyzed the actions that HCWs performed with patients' EHRs to measure worker-worker connections. There are six types of HCW actions, including condition-related (eg, assigning a diagnosis), procedure-related (eg, intubation), medication-related (eg, prescriptions), note-related (eg, writing progress notes), order-related (eg, ordering laboratory tests), and measurement-related (eg, measuring respiratory rate) actions.

Research has shown that a 1-day window is enough to capture the meaningful, collaborative relationships among HCWs [20-23]. Therefore, we assumed that there was a connection between two HCWs who interacted with the same patient's EHR on the same day. We built a network in which the nodes represented HCWs and the edges indicated the number of days that two HCWs performed actions on the EHRs of the same patients. We built one network for critically ill patients with COVID-19 and another for patients without COVID-19.

The nodes in the COVID-19 and non–COVID-19 networks were defined as follows:

$$Z_{COVID\text{-}19} = \{z_1, z_2, \ldots, z_p\} \quad (1)$$

$$Z_{Non\text{–}COVID\text{-}19} = \{z'_1, z'_2, \ldots, z'_q\} \quad (2)$$

To better interpret the networks, we used an HCW's specialty (eg, respiratory care) and type (eg, respiratory therapist) to label each node. We combined these factors to define expertise (ie, "specialty: type"; eg, respiratory care: respiratory therapist). Expertise in the COVID-19 and non–COVID-19 networks were defined as follows:

$$EXP_{COVID\text{-}19} = \{exp_1, exp_2, \ldots, exp_a\} \quad (3)$$

$$EXP_{Non\text{–}COVID\text{-}19} = \{exp'_1, exp'_2, \ldots, exp'_b\} \quad (4)$$

In equations 1-4, Z and EXP were used to describe the composition of COVID-19 or non–COVID-19 networks.

In each network, we used two sociometrics—eigenvector centrality and betweenness centrality—to quantify an HCW's core and betweenness status in the network, respectively. We used Gephi (ie, an open-source network analysis and visualization software package) [25] to calculate eigencentrality and betweenness centrality values.

Eigencentrality characterizes the degree to which an HCW is densely connected to other HCWs who are also densely connected with other HCWs. A high-eigencentrality HCW is likely to be a core person who actively works with other HCWs when performing actions on EHRs. An example HCW network with eigencentrality values is shown in Multimedia Appendix 2.

The betweenness centrality of an HCW refers to the number of shortest paths between two other HCWs that pass through the HCW in question. An HCW with a broad skill set who cares for a wide spectrum of patients could frequently be in a high-betweenness position. An example HCW network with betweenness centrality values is shown in Multimedia Appendix 2.

Eigencentrality and Betweenness in COVID-19 and Non–COVID-19 Networks

We investigated whether differences in the eigencentrality and betweenness of COVID-19 and non–COVID-19 critical care structures were statistically significant at the network and expertise levels. The network-level comparison was conducted by assessing the network as a whole (ie, COVID-19 vs non–COVID-19 networks), while an expertise-level comparison was conducted for each expertise (eg, internal medicine physicians). Since eigencentrality and betweenness are not Gaussian distributed, we conducted a Mann-Whitney U test with a significance level of α=.05. The tests for expertise included at least 8 HCWs and involved Bonferroni correction to account for multiple hypotheses.

Patient Staffing Intensity in COVID-19 and Non–COVID-19 Settings

We defined the set of inpatient stays in COVID-19 and non–COVID-19 settings as follows:

$$S_{COVID\text{-}19} = \{s_1, s_2, \ldots, s_m\} \quad (5)$$

$$S_{Non\text{–}COVID\text{-}19} = \{s'_1, s'_2, \ldots, s'_n\} \quad (6)$$

Since each inpatient stay (ie, s_i) can last for more than 1 day, we defined the jth day of a stay as $s_{i,j}$ (ie, $1 \le j \le l_i$); l_i represents the last day of a patient's hospital stay (ie, s_i). For $s_{i,j}$, we calculated the number of HCWs (ie, $Ns_{i,j,}exp_k$) in each expertise category (ie, exp_k) who interacted with the EHRs of patient i on day j. For each inpatient stay (ie, s_i), we calculated the average number of HCWs in each expertise category (ie, exp_k) who interacted with the EHRs of the same patient on each day, as follows:

In equation 7, los_i refers to the length of hospital stay (ie, the total number of hours between the start and end times of an inpatient stay divided by 24 hours). Since each inpatient stay may start and end at different times of the day,los_i may be different from l_i. Daily patient staffing intensity was defined as an expertise-level value (ie, ).

To learn about the differences in the daily patient staffing intensities of COVID-19 and non–COVID-19 critical care, we conducted a set of tests. Specifically, for each investigated expertise (eg, internal medicine nurse practitioners), we tested whether critically ill patients with COVID-19 required a significantly higher daily patient staffing intensity than critically ill patients without COVID-19. We focused on the 20 expertise categories with the highest mean daily staffing intensity values in COVID-19 and non–COVID-19 critical care and used the Mann-Whitney U test, which had a Bonferroni-corrected significance level of .05.

We also assessed the differences in the overall patient staffing intensities of COVID-19 and non–COVID-19 critical care in terms of the number of HCWs who were involved in the management of a patient. Overall staffing intensity was defined as the number of HCWs who interacted with the EHRs of a

patient with or without COVID-19. Our hypothesis was as follows: critically ill patients with COVID-19 require a significantly higher overall staffing intensity than critically ill patients without COVID-19. We tested this hypothesis by using the Mann-Whitney U test, which had a Bonferroni-corrected significance level of .05.

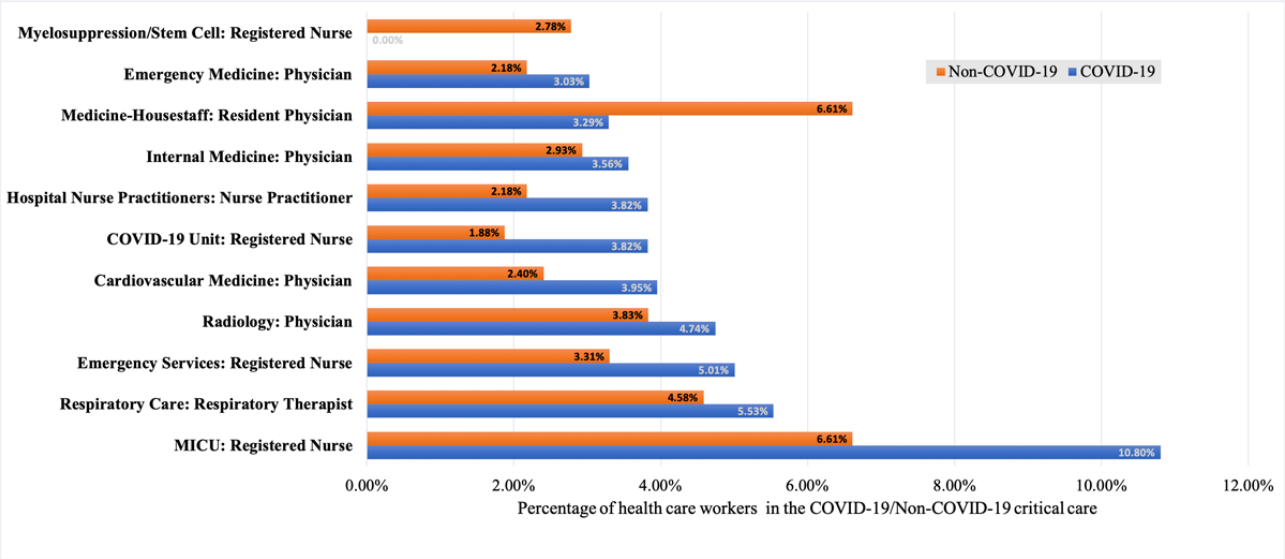
Results

HCW Characteristics

The number of HCWs, types of HCWs, specialties, and expertise categories in COVID-19 and non–COVID-19 critical care was 759 and 1331, 24 and 24, 92 and 128, and 133 and 207, respectively. These values indicated that patients without COVID-19 required more expertise, specialties, and HCWs. A possible reason for this is that critically ill patients without COVID-19 were admitted to the MICU for a wide range of major conditions. With regard to patient-level values, COVID-19 and non–COVID-19 critical care consisted of 79.5 and 88.2 HCWs, 9.8 and 10.6 types of HCWs, 23.0 and 27.1 departments, and 29.2 and 34.0 expertise categories, respectively. The patient-level values for COVID-19 and non–COVID-19 critical care were highly similar.

Figure 1 illustrates the union of the 10 COVID-19 and non–COVID-19 critical care expertise categories with the largest proportion of HCWs. It can be seen that, aside from residents and registered nurses with myelosuppression expertise, the COVID-19 setting had higher percentages of different types of HCWs than the non–COVID-19 setting. These results demonstrate how the Vanderbilt University Medical Center assigned full-time, nontrainee HCWs to the task of managing critically ill patients with COVID-19 and reduced the number of residents during the COVID-19 pandemic.

Figure 1. The expertise categories with the largest number of health care workers in the COVID-19 and non–COVID-19 settings. There are 11 expertise categories shown, which correspond to the union of the top 10 expertise categories in COVID-19 (ie, excluding nurses with myelosuppression expertise) and non–COVID-19 (ie, excluding COVID-19 unit nurses) critical care. Each expertise is reported in the following format: "specialty: health care worker type." MICU: medical intensive care unit.

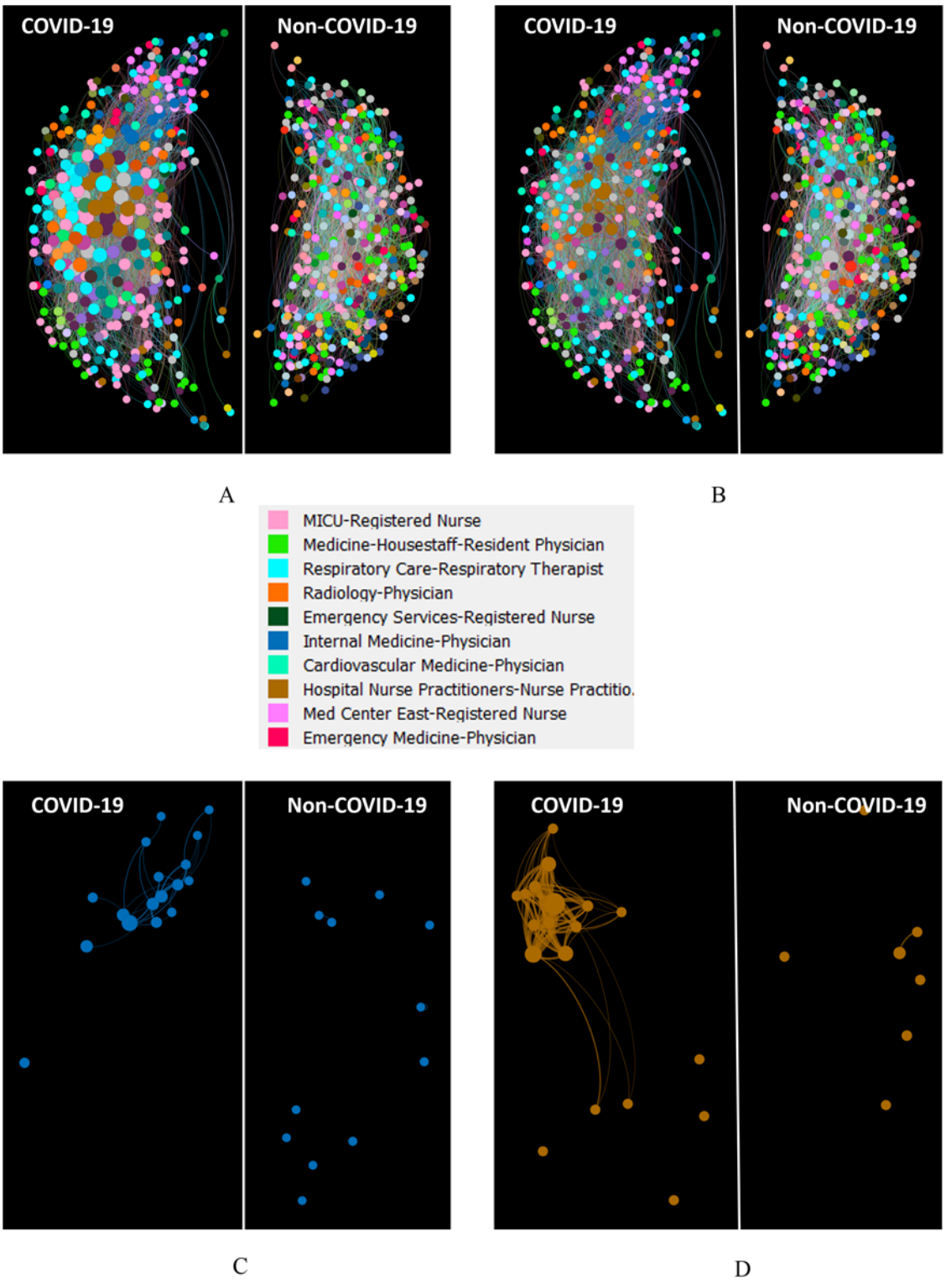


There were no registered nurses with myelosuppression expertise in the COVID-19 setting. Upon further analysis, we found 4 patients with COVID-19 and cancer, but none were in need of invasive intervention at the time of their care.

Eigencentrality and Betweenness

Figure 2 presents the HCW networks in COVID-19 and non–COVID-19 critical care from eigencentrality and betweenness perspectives. In the figure, it can be seen that the majority of HCWs in the COVID-19 network are larger in size (ie, higher eigencentrality) than those in the non–COVID-19 network. This indicates that HCWs are much more highly and densely connected in the COVID-19 network than those in the non–COVID-19 network. We performed a test to measure the differences in eigencentrality between the COVID-19 and non–COVID-19 networks. The results indicated that the two networks had significantly different median eigencentrality values (COVID-19: 0.096; non–COVID-19: 0.057; $P<.001$).

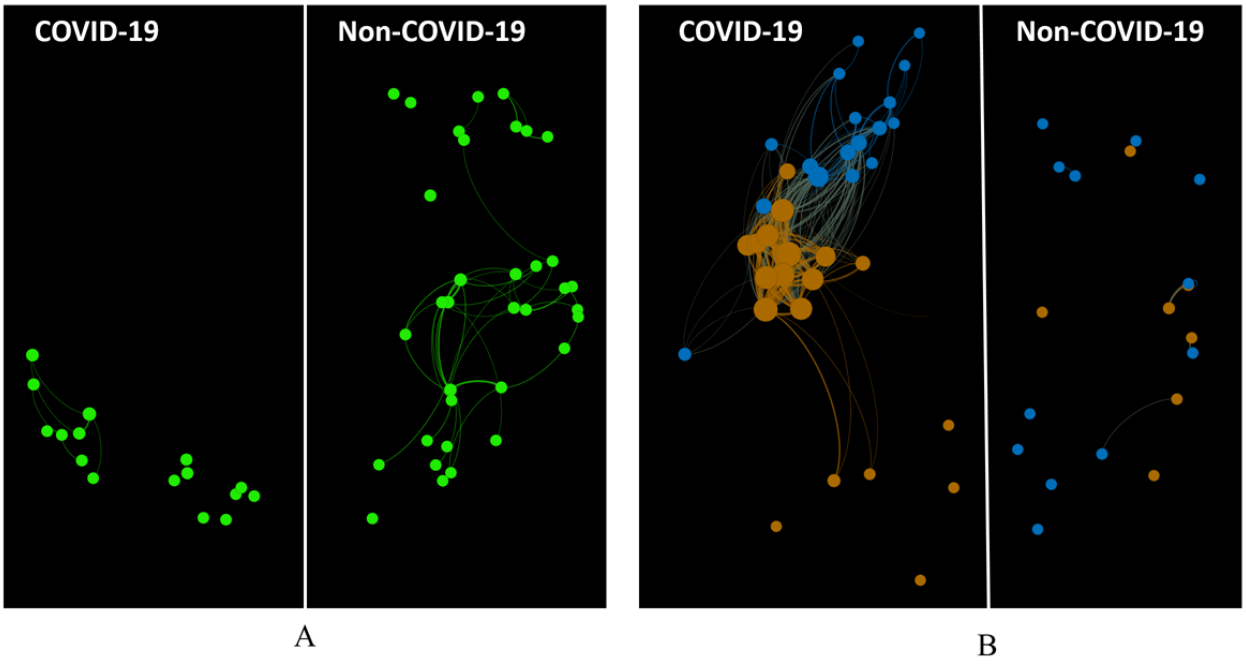
Figure 2. A depiction of the health care worker eigencentrality (A) and betweenness (B) in COVID-19 and non–COVID-19 networks. C and D show the subnetworks of internal medicine physicians and nurse practitioners in the COVID-19 and non–COVID-19 networks, respectively. In A and C, eigencentrality directly correlated with the size of the corresponding node. In B and D, betweenness centrality directly correlated with the size of the corresponding node. The legend in the figure shows the 10 expertise categories with the largest number of health care workers in the combined network (ie, both the COVID-19 and non–COVID-19 networks). MICU: medical intensive care unit.



After removing expertise categories that had less than 8 HCWs, we performed pairwise tests on the remaining 12 expertise categories. The results of these tests are provided in Table S1 in Multimedia Appendix 3. There were several notable findings. First, we observed that internal medicine physicians in the COVID-19 network had higher eigencentrality values than those in the non–COVID-19 network (P=.001). Figure 2 also shows the subnetworks of internal medicine physicians in the COVID-19 and non–COVID-19 networks from an eigencentrality perspective. From the figure, it can be seen that internal medicine physicians in the COVID-19 network were connected with each other, while those in the non–COVID-19

network were separated. Second, the resident physicians in the non–COVID-19 network had higher eigencentrality values than those in the COVID-19 network (P=.002). Figure 3 shows that there were many residents across the entire non–COVID-19 network. However, the number of residents was much smaller in the COVID-19 network.

Figure 3. (A) The subnetworks of resident physicians in the COVID-19 and non–COVID-19 networks. (B) The subnetworks of internal medicine physicians and nurse practitioners in the COVID-19 and non–COVID-19 networks. Eigencentrality directly correlated with the size of the corresponding node in both A and B.



There were no significant differences in the betweenness of the two networks (COVID-19 network: median 0.002; non–COVID-19 network: median 0.003; P=.22). However, nurse practitioners in the COVID-19 network had significantly higher betweenness values than those in the non–COVID-19 network (P<.001). The complete set of test results for expertise categories with nonsignificantly different betweenness values is provided in Table S2 in Multimedia Appendix 3. Figure 2 also shows the subnetworks of nurse practitioners in the COVID-19 and non–COVID-19 networks from a betweenness perspective. From the figure, it can be seen that nurse practitioners had a larger number of connections in the COVID-19 network than those in the non–COVID-19 network. In Figure 2, it can also be seen that nurse practitioners are in the central part of the COVID-19 network and serve as connective bridges between other HCWs. Given that betweenness reflects an HCW's access to a wide spectrum of patients, a nurse practitioner in the COVID-19 collaboration structure can build connections among HCWs who are not directly connected. Internal medicine physicians and nurse practitioners were the core of the COVID-19 network. As shown in Figure 3, these two types of HCWs had a larger number of connections in the COVID-19 care setting than those in the non–COVID-19 setting.

Differences in Patient Staffing Intensity

In total, 41,903 (mean 1103) actions were performed with the EHRs of patients with COVID-19 and 44,131 (mean 1161) actions were performed with the EHRs of patients without COVID-19. There were no statistically significant differences in the number of actions performed with the EHRs of patients with and without COVID-19 (P=.32). The differences in the number of expertise categories and HCWs who performed actions on the EHRs of critically ill patients with and without COVID-19 were also not significant (expertise categories: P=.08; HCWs: P=.19). The complete set of results is provided in Table S3 in Multimedia Appendix 4.

Table 2 shows the differences in the patient staffing intensities of COVID-19 and non–COVID-19 critical care for each expertise category. The union of 20 COVID-19 and 20 non–COVID-19 expertise categories yielded 24 categories.

Table 2. Differences in the daily average patient staffing intensities of COVID-19 and non–COVID-19 critical care. A Bonferroni-corrected *P* value of .002 was used as the null hypothesis rejection threshold.

Expertise category	Staffing intensity, mean (SD)		Staffing intensity, median (IQR)		*P* value
	COVID-19 critical care	Non–COVID-19 critical care	COVID-19 critical care	Non–COVID-19 critical care	
COVID-19 categories					
Hospital nurse practitioners: nurse practitioner	1.81 (0.19)	0.27 (0.10)	1.76 (1.37)	0 (0.18)	<.001
Medical Center East[a]: registered Nurse	0.78 (0.08)	0.21 (0.07)	0.69 (0.70)	0 (0.02)	<.001
Cardiovascular intensive care unit: registered Nurse	0.26 (0.05)	0.01 (0.01)	0.18 (0.37)	0 (0)	<.001
Internal medicine: nurse practitioner	0.33 (0.05)	0.06 (0.02)	0.23 (0.30)	0 (0.05)	<.001
Surgical intensive care unit: registered nurse	0.17 (0.04)	0.03 (0.02)	0.09 (0.29)	0 (0)	.002
Non–COVID-19 categories					
Medicine house staff: resident physician	0.18 (0.06)	1.56 (0.15)	0 (0.07)	1.62 (1.32)	<.001
Emergency medicine: resident physician	0.06 (0.03)	0.27 (0.05)	0 (0.05)	0.16 (0.30)	<.001
Categories that were not statistically significant					
Medical Center East[a]: technician	0.23 (0.03)	0.10 (0.04)	0.22 (0.19)	0 (0)	.007
Hematology oncology: physician	$3.54\ 10^{-3}$ ($2.87\ 10^{-3}$)	0.22 (0.08)	0 (0)	0 (0.06)	.01
Emergency medicine: physician	0.27 (0.06)	0.53 (0.10)	0.11 (0.43)	0.30 (0.92)	.02
Pharmacy inpatient (central): pharmacist	0.12 (0.03)	0.21 (0.03)	0.06 (0.14)	0.15 (0.13)	.03
Pharmacy inpatient (evening): pharmacist	0.16 (0.03)	0.24 (0.03)	0.12 (0.28)	0.21 (0.23)	.03
Infectious disease: physician	0.05 (0.03)	0.18 (0.05)	0 (0)	0 (0.20)	.05
Internal medicine: physician	1.14 (0.13)	0.60 (0.11)	0.92 (1.12)	0.53 (0.90)	.002
Medical intensive care unit: registered nurse	1.19 (0.13)	0.95 (0.10)	1.03 (0.91)	0.85 (0.78)	.13
Radiology: physician	0.30 (0.04)	0.38 (0.04)	0.27 (0.34)	0.37 (0.42)	.15
Nephrology: physician	0.14 (0.05)	0.26 (0.10)	0 (0)	0 (0.05)	.21
Allergy/pulmonary: physician	0.80 (0.11)	0.94 (0.12)	0.65 (0.82)	0.88 (0.73)	.21
Pharmacy inpatient operations manager: pharmacist	0.14 (0.02)	0.21 (0.04)	0.11 (0.13)	0.14 (0.24)	.22
Pharmacy inpatient satellite operating room: pharmacist	0.20 (0.04)	0.15 (0.03)	0.11 (0.38)	0.11 (0.23)	.28
Respiratory care: respiratory therapist	0.74 (0.11)	0.89 (0.14)	0.60 (0.87)	0.80 (1.32)	.30
Emergency services: registered nurse	0.15 (0.05)	0.17 (0.05)	0.07 (0.18)	0.09 (0.18)	.41
Pharmacy inpatient (evening): pharmacy technician	0.26 (0.05)	0.28 (0.05)	0.16 (0.36)	0.19 (0.40)	.43
Cardiovascular medicine: physician	0.21 (0.04)	0.18 (0.03)	0.15 (0.37)	0.16 (0.15)	.43

[a]Medical Center East is the building where we created the COVID-19 unit. Before the creation of the COVID-19 unit, nurses in this building cared for critically ill patients without COVID-19.

There was a larger number of internal medicine nurse practitioners, cardiovascular ICU registered nurses, and surgical ICU registered nurses who performed daily actions on the EHRs of critically ill patients with COVID-19 compared to the number of those who performed daily actions on the EHRs of patients without COVID-19. In contrast, the EHRs of patients without COVID-19 were managed by a larger number of resident physicians (ie, those with medicine and emergency medicine expertise). These differences were statistically significant (Table 2).

We also found that expertise categories were not statistically different in terms of daily patient staffing intensity. Such categories included radiology physicians, nephrology physicians, pulmonary/allergy physicians, emergency medicine physicians, MICU registered nurses, and respiratory therapists.

Discussion

Principal Findings

There are no universal guidelines for HCW staffing in ICUs. To date, ICU staffing studies have focused on organization models (eg, open, closed, and hybrid models), and few have examined collaborations among HCWs. In this study, we used a novel method for learning about collaborations among HCWs and building corresponding networks. We measured eigencentrality and betweenness centrality to quantify the core and betweenness statuses of HCWs and identify several significant differences between the COVID-19 and non–COVID-19 network structures. Differences in the collaboration structures between the two networks mirrored those in intentional strategic planning structures across the health care system. For instance, there was a significant difference (P<.001) in the number of resident physicians between the COVID-19 and non–COVID-19 structures because our medical center assigned full-time, nontrainee HCWs to the management of critically ill patients with COVID-19. This mirrors resident protection strategies that were implemented during the outset of the COVID-19 pandemic by the National Graduate Medical Education. Figure 3 shows the subnetworks of resident physicians in the COVID-19 and non–COVID-19 networks. It can be seen that the non–COVID-19 network has a larger resident network than the COVID-19 network, and the connections between residents are more complex than those in the COVID-19 network. This suggests that resident physicians are highly active with respect to the management of critically ill patients in a non–COVID-19 setting.

Beyond collecting data on basic strategic planning methods (ie, the reduction of the number of residents) in the management of critically ill patients with COVID-19, we also learned about the aspects of collaboration structures that are important for the management of teamwork but are not explicitly documented in existing ICU staffing plans. We found that internal medicine physicians and nurse practitioners in the COVID-19 collaboration structure were more active (ie, high eigencentrality or betweenness) than those in the non–COVID-19 collaboration structure. As shown in Figure 3, internal medicine physicians and nurse practitioners connected more frequently in the COVID-19 network than those in the non–COVID-19 network. This phenomenon suggests that they are core members in collaborations that relate to the management of critically ill patients with COVID-19.

Combining knowledge on connections among HCWs with their eigencentrality and betweenness values in the collaboration network can assist HCOs with designing and developing more specific staffing strategies, which can potentially improve care quality and patient outcomes. The network analysis methods and team structures that are depicted in our retrospective study can be used in a prospective setting. Our COVID-19 and non–COVID-19 networks can be used to identify the characteristics of a newly established or modified team. For instance, if a COVID-19 ICU has plans for creating a team to care for the increasing number of patients, the eigencentrality and betweenness centrality of each HCW and the HCW relationships that we learned about in our COVID-19 network can be used as evidence for identifying the characteristics of the newly created team. The team creators can evaluate the leadership (ie, eigencentrality), robustness (ie, betweenness), and familiarity (ie, the strength of the relationships between HCWs) of the newly established team. They can also dynamically add or remove an HCW from the created team and measure changes in leadership, robustness, and familiarity, which will help team creators with finding their desired team.

The Scope of This Study and Its Limitations

In this study, we did not investigate temporal networks or team dynamics, which are essential to HCOs that monitor and manage team dynamics. However, researchers can use the network analysis methodologies that were developed in our study to identify temporal networks, such as daily, weekly, or monthly networks. They can also use the sociometrics that we developed to quantify changes in temporal network structures. For instance, HCOs can use our network analysis methodologies to temporally measure the relationships among internal medicine physicians, nurse practitioners, and residents; and quantify the weekly, monthly, and yearly changes in these relationships.

There are several limitations in this study that should be recognized. These limitations serve as opportunities for further investigation. First, this study was based on a small number of critically ill patients with COVID-19. Although our sample had sufficient power for analyzing the differences in the eigencentrality, betweenness, and patient staffing intensities for several types of HCWs in COVID-19 and non–COVID-19 critical care, a larger volume of data is needed to obtain statistically meaningful results. Second, comorbidities could have impacted team structures; however, matching the comorbidities between patients with and without COVID-19 can lead to certain risks. According to our observations, matching comorbidities between the two cohorts will considerably enlarge the study window (ie, >3 years) for the non–COVID-19 cohort. However, non–COVID-19 care teams can drastically change over time, making the study of the non–COVID-19 team structures less meaningful. Therefore, in this study, we focused on the most important confounding factors (ie, age, gender, and the length of stay) that characterize team efficacy and may impact team structures. Additionally, our study's primary focus was to learn about team structures in

COVID-19 and medical ICUs, which allowed for some degree of variance in comorbidities. Third, the characteristics of the COVID-19 structures that we learned about during this single-center study could provide HCOs with reference data for assessing their own COVID-19 ICU structures. However, our medical center is an institution that intentionally developed a nurse practitioner–centered organizational structure. This should be considered when interpreting our results and findings. To learn about general COVID-19 ICU collaboration structures, researchers need to conduct analyses that account for multiple HCOs. Fourth, there was a lack of standard terminology for characterizing HCO departments and the roles of HCWs. Although there are taxonomies for describing clinician specialties [26-28], these tend to neglect the nonphysicians who play vital roles in the management of patients. It is clear that common data models for department names and HCW types would improve the quality of our study and assist other institutions with using our methodology. Fifth, we assumed that two HCWs would have a connection when they performed actions on the EHRs of patients. Although such an assumption can help with identifying collaboration relationships between HCWs, it may also have resulted in the identification of many spurious relationships.

Conclusion

HCOs have been planning and refining their staffing strategies to provide more efficient and effective care to patients with COVID-19. However, there are few efficient methodologies for assessing the execution of collaboration structures in practice, especially those for assessing the cross-disciplinary connections among HCWs. In this study, we demonstrated how data on the use of a large academic medical center's EHR system could be used to learn about the collaboration structures in COVID-19 critical care (ie, through network analysis methodologies). HCOs can use our network analysis approaches and data on eigencentrality, betweenness, and patient staffing intensities to characterize HCW roles in collaboration networks during the COVID-19 pandemic or future events. Furthermore, research on how HCWs are connected has created an opportunity for studying the relationships among team structures, care quality, and patient safety.

Acknowledgments

This study was partially supported by the National Library of Medicine of the National Institutes of Health (award number: R01LM012854). The content of this study is solely the authors' responsibility; it does not necessarily represent the National Institutes of Health's official views. The data sets that were used for the described analyses were obtained from Vanderbilt University Medical Center's Research Derivative, which is supported by institutional funds and Vanderbilt Clinical and Translational Science Awards (grant number: ULTR000445) from the National Center for Advancing Translational Sciences and the National Institutes of Health. The Vanderbilt Institutional Review Board reviewed and approved this study (approval number: 200792).

The data sets that were generated and analyzed during this study are not publicly available because they include patients' private information. However, the data sets can be obtained from the corresponding author upon reasonable request.

Authors' Contributions

YC conceived the study idea, collected the data, analyzed the data, designed the methods, designed the experiment, evaluated and interpreted the results of the experiment, and wrote the manuscript. CY conceived the study idea, analyzed the data, designed the methods, designed the experiment, evaluated and interpreted the results of the experiment, and revised the manuscript. XZ analyzed the data, evaluated and interpreted the results of the experiment, and revised the manuscript. CG analyzed the data, evaluated and interpreted the results of the experiment, and revised the manuscript. EW, JC, DF, YG, MP, and BM evaluated and interpreted the results of the experiment and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Distributions of the COVID-19 and non–COVID-19 groups' propensity scores.
[DOCX File , 258 KB - humanfactors_v8i1e25724_app1.docx]

Multimedia Appendix 2
Examples that illustrate eigencentrality and betweenness centrality.
[DOCX File , 71 KB - humanfactors_v8i1e25724_app2.docx]

Multimedia Appendix 3
Differences in the eigencentrality and betweenness of COVID-19 and non–COVID-19 structures.
[DOCX File , 37 KB - humanfactors_v8i1e25724_app3.docx]

Multimedia Appendix 4
Differences in overall patient staffing intensities between the COVID-19 and non–COVID-19 structures.
[DOCX File , 31 KB - humanfactors_v8i1e25724_app4.docx]

References

1. COVID data tracker weekly report. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html [accessed 2021-02-25]
2. Li R, Rivers C, Tan Q, Murray MB, Toner E, Lipsitch M. Estimated demand for US hospital inpatient and intensive care unit beds for patients with COVID-19 based on comparisons with Wuhan and Guangzhou, China. JAMA Netw Open 2020 May 01;3(5):e208297 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.8297] [Medline: 32374400]
3. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et al. Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team. 2020 Mar 16. URL: https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf [accessed 2021-02-25]
4. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med 2020 Apr 30;382(18):1708-1720 [FREE Full text] [doi: 10.1056/NEJMoa2002032]
5. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of Covid-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/NEJMsb2005114] [Medline: 32202722]
6. Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. Ann Intern Med 2020 Jul 07;173(1):21-28 [FREE Full text] [doi: 10.7326/M20-1260] [Medline: 32259197]
7. Chopra V, Toner E, Waldhorn R, Washer L. How should U.S. hospitals prepare for coronavirus disease 2019 (COVID-19)? Ann Intern Med 2020 May 05;172(9):621-622 [FREE Full text] [doi: 10.7326/M20-0907] [Medline: 32160273]
8. Anstey DE, Givens R, Clerkin K, Fried J, Kalcheva N, Kumaraiah D, et al. The cardiac intensive care unit and the cardiac intensivist during the COVID-19 surge in New York City. Am Heart J 2020 Sep;227:74-81 [FREE Full text] [doi: 10.1016/j.ahj.2020.06.018] [Medline: 32682106]
9. Harkouk H, Jacob C, Fletcher D. Urgent development of an anaesthesiology-based intensive care unit for critical COVID-19 infected patients. Anaesth Crit Care Pain Med 2020 Jun;39(3):359-360 [FREE Full text] [doi: 10.1016/j.accpm.2020.04.011] [Medline: 32389563]
10. Kotfis K, Roberson SW, Wilson JE, Dabrowski W, Pun BT, Ely EW. COVID-19: ICU delirium management during SARS-CoV-2 pandemic. Crit Care 2020 Apr 28;24(1):176 [FREE Full text] [doi: 10.1186/s13054-020-02882-x] [Medline: 32345343]
11. Donovan AL, Aldrich JM, Gross AK, Barchas DM, Thornton KC, Schell-Chaple HM, University of California, San Francisco Critical Care Innovations Group. Interprofessional care and teamwork in the ICU. Crit Care Med 2018 Jun;46(6):980-990. [doi: 10.1097/CCM.0000000000003067] [Medline: 29521716]
12. Wilkes MS, Marcin JP, Ritter LA, Pruitt S. Organizational and teamwork factors of tele-intensive care units. Am J Crit Care 2016 Sep;25(5):431-439. [doi: 10.4037/ajcc2016357] [Medline: 27587424]
13. Rosen MA, DiazGranados D, Dietz AS, Benishek LE, Thompson D, Pronovost PJ, et al. Teamwork in healthcare: Key discoveries enabling safer, high-quality care. Am Psychol 2018;73(4):433-450 [FREE Full text] [doi: 10.1037/amp0000298] [Medline: 29792459]
14. Sandoff M, Nilsson K. How staff experience teamwork challenges in a new organizational structure. Team Performance Management 2016 Oct 10;22(7/8):415-427. [doi: 10.1108/TPM-05-2016-0021]
15. Bartsch S, Weber E, Büttgen M, Huber A. Leadership matters in crisis-induced digital transformation: how to lead service employees effectively during the COVID-19 pandemic. Journal of Service Management 2020 Aug 12;32(1):71-85 [FREE Full text] [doi: 10.1108/josm-05-2020-0160]
16. Jamoom EW, Yang N, Hing E. Adoption of certified electronic health record systems and electronic information sharing in physician offices: United States, 2013 and 2014. NCHS Data Brief 2016 Jan(236):1-8 [FREE Full text] [Medline: 26828707]
17. Holmgren AJ, Adler-Milstein J. Health information exchange in US hospitals: The current landscape and a path to improved information sharing. J Hosp Med 2017 Mar;12(3):193-198. [doi: 10.12788/jhm.2704] [Medline: 28272599]
18. Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health record systems among US non-federal acute care hospitals?2015. The Office of the National Coordinator for Health Information Technology. 2016 May. URL: https://dashboard.healthit.gov/evaluations/data-briefs/non-federal-acute-care-hospital-ehr-adoption-2008-2015.php [accessed 2021-02-25]
19. Adler-Milstein J, Holmgren AJ, Kralovec P, Worzala C, Searcy T, Patel V. Electronic health record adoption in US hospitals: the emergence of a digital "advanced use" divide. J Am Med Inform Assoc 2017 Nov 01;24(6):1142-1148 [FREE Full text] [doi: 10.1093/jamia/ocx080] [Medline: 29016973]

20. Chen Y, Lorenzi NM, Sandberg WS, Wolgast K, Malin BA. Identifying collaborative care teams through electronic medical record utilization patterns. J Am Med Inform Assoc 2017 Apr 01;24(e1):e111-e120 [FREE Full text] [doi: 10.1093/jamia/ocw124] [Medline: 27570217]
21. Chen Y, Lehmann CU, Hatch LD, Schremp E, Malin BA, France DJ. Modeling care team structures in the neonatal intensive care unit through network analysis of EHR audit logs. Methods Inf Med 2019 Nov;58(4-05):109-123 [FREE Full text] [doi: 10.1055/s-0040-1702237] [Medline: 32170716]
22. Kim C, Lehmann CU, Hatch D, Schildcrout JS, France DJ, Chen Y. Provider networks in the neonatal intensive care unit associate with length of stay. IEEE Conf Collab Internet Comput 2019 Dec;2019:127-134 [FREE Full text] [doi: 10.1109/CIC48465.2019.00024] [Medline: 32637942]
23. Chen Y, Patel MB, McNaughton CD, Malin BA. Interaction patterns of trauma providers are associated with length of stay. J Am Med Inform Assoc 2018 Jul 01;25(7):790-799 [FREE Full text] [doi: 10.1093/jamia/ocy009] [Medline: 29481625]
24. Chen Y, Yan C, Patel MB. Network analysis subtleties in ICU structures and outcomes. Am J Respir Crit Care Med 2020 Dec 01;202(11):1606-1607 [FREE Full text] [doi: 10.1164/rccm.202008-3114LE] [Medline: 32931298]
25. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Proceedings of the Third international AAAI conference on weblogs and social media. 2009 Presented at: The Third international AAAI conference on weblogs and social media; May 17-20, 2009; San Jose, California URL: https://gephi.org/publications/gephi-bastian-feb09.pdf [doi: 10.1007/978-1-4614-6170-8_299]
26. Turer RW, Jones I, Rosenbloom ST, Slovis C, Ward MJ. Electronic personal protective equipment: A strategy to protect emergency department providers in the age of COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):967-971 [FREE Full text] [doi: 10.1093/jamia/ocaa048] [Medline: 32240303]
27. Bindman AB. Using the national provider identifier for health care workforce evaluation. Medicare Medicaid Res Rev 2013 Jul 30;3(3):mmrr.003.03.b03 [FREE Full text] [doi: 10.5600/mmrr.003.03.b03] [Medline: 24753977]
28. Gutowski J. New taxonomy codes and the IBCLC. Clin Lactation 2011;2(3):34 [FREE Full text] [doi: 10.1891/215805311807010459]

Abbreviations

EHR: electronic health record
HCO: health care organization
HCW: health care worker
ICU: intensive care unit
MICU: medical intensive care unit

Edited by A Kushniruk; submitted 13.11.20; peer-reviewed by A Mohammed, J Walsh; comments to author 23.12.20; revised version received 12.01.21; accepted 22.02.21; published 08.03.21.

Please cite as:
Yan C, Zhang X, Gao C, Wilfong E, Casey J, France D, Gong Y, Patel M, Malin B, Chen Y
Collaboration Structures in COVID-19 Critical Care: Retrospective Network Analysis Study
JMIR Hum Factors 2021;8(1):e25724
URL: https://humanfactors.jmir.org/2021/1/e25724
doi:10.2196/25724
PMID:33621187

Original Paper

A Clinical Journey Mobile Health App for Perioperative Patients: Cross-sectional Study

Stijn J Willems[1], MSc; Michel W Coppieters[1,2], PhD; Yvette Pronk[3], MSc; Miranda J F Diks[3], MD; Klaas W A P van der Heijden[3], MD; Servan Rooker[3,4], MD, PhD; Gwendolyne G M Scholten-Peeters[1], PhD

[1]Department of Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
[2]Menzies Health Institute Queensland, Griffith University, Brisbane & Gold Coast, Australia
[3]Department of Orthopaedics and Research, Kliniek ViaSana, Mill, Netherlands
[4]Department of Epidemiology and Social Medicine, University of Antwerp, Antwerp, Belgium

Corresponding Author:
Gwendolyne G M Scholten-Peeters, PhD
Department of Human Movement Sciences
Faculty of Behavioural and Movement Sciences
Vrije Universiteit Amsterdam
Van der Boechorststraat 9
Amsterdam
Netherlands
Phone: 31 20 59 88557
Email: g.g.m.scholten-peeters@vu.nl

Abstract

Background: Mobile eHealth apps are important tools in personal health care management. The Patient Journey app was developed to inform patients with musculoskeletal disorders during their perioperative period. The app contains timely information, video exercises, and functional tasks. Although the Patient Journey app and other health apps are widely used, little research is available on how patients appreciate these apps.

Objective: The primary aim of this study was to evaluate the user-friendliness of the Patient Journey app in terms of its usability and the attitudes of users toward the app. The secondary aim was to evaluate positive and negative user experiences.

Methods: A web-based questionnaire was sent to 2114 patients scheduled for surgery for a musculoskeletal disorder. Primary outcomes were usability (measured with the System Usability Scale) and user attitudes regarding the Patient Journey app (assessed with the second part of the eHealth Impact Questionnaire). The secondary outcomes were evaluated with multiple choice questions and open-ended questions, which were analyzed via inductive thematic content analyses.

Results: Of the 940 patients who responded, 526 used the Patient Journey app. The usability of the app was high (System Usability Scale: median 85.0, IQR 72.5-92.5), and users had a positive attitude toward the Information and Presentation provided via the app (eHealth Impact Questionnaire: median 78.0, IQR 68.8-84.4). The app did not adequately improve the users' confidence in discussing health with others (eHealth Impact Questionnaire: median 63.9, IQR 50.0-75.0) or motivation to manage health (eHealth Impact Questionnaire: median 61.1, IQR 55.6-72.2). Three core themes emerged regarding positive and negative user experiences: (1) content and information, (2) expectations and experiences, and (3) technical performance. Users experienced timely information and instructions positively and found that the app prepared and guided them optimally through the perioperative period. Negative user experiences were overly optimistic information, scarcely presented information about pain (medication), lack of reference data, insufficient information regarding clinical course deviations and complications, and lack of interaction with clinicians.

Conclusions: The Patient Journey app is a usable, informative, and presentable tool to inform patients with musculoskeletal disorders during their perioperative period. The qualitative analyses identified aspects that can further improve the user experiences of the app.

(JMIR Hum Factors 2021;8(1):e20694) doi:10.2196/20694

KEYWORDS

eHealth; mHealth; applications; musculoskeletal; user-friendliness; rehabilitation; usability; patient education; technology; disability; feasibility; adherence

Introduction

eHealth and mobile health (mHealth) tools have the potential to enhance the quality of health care and to reduce health care costs [1]. Consequently, the use of eHealth and mHealth can play an important role in supporting personal health management by encouraging healthy behavior and improving adherence and self-management [2,3]. mHealth can have additional value because only a limited amount of medical information can correctly be remembered after a consultation, and mHealth apps can be used at any time and any place [4-6]. This can enhance information recall and adherence to health instructions [5,7,8]. Furthermore, recent research shows that education provided to patients through their smartphone may improve their levels of knowledge, medication or treatment adherence, satisfaction, and clinical outcomes, as well as having a positive effect on health care economics [9].

Previous research showed that the use of mHealth apps is well appreciated by users during the perioperative period in different health care settings [10,11]. Reported advantages are the patient's sense of being looked after, enhancement of patient-centered care, cost-effectiveness, and the increased efficiency of health care services [10,11]. However, to date, the user experiences of health care apps for the perioperative guidance of musculoskeletal surgeries have not yet been evaluated.

Based on these advantages, we evaluated the user experience of a widely used mHealth app called the Patient Journey app for patients with musculoskeletal disorders. The app provides timely information, exercises tailored to the condition and recovery, and functional tasks. The app was developed with the assumption that it addresses the patients' needs better at specific time points and improves self-management compared to traditionally provided information.

Even though the app is widely used by over 100 hospitals and clinics in more than 20 countries, evidence about how patients appreciate this app is not yet available. Before an effectiveness study can be performed, the user-friendliness of the app needs to be assessed. Therefore, the primary aim of this study was to evaluate the user-friendliness in terms of usability and the attitudes of users toward the app. The secondary aim was to explore positive and negative user experiences.

Methods

Study Design

This was a cross-sectional user-friendliness study using digital surveys. The study was approved by the local medical ethics committee of Vrije Universiteit Amsterdam (VCWE-2017-005). All patients provided digital informed consent prior to participating in the study.

Recruitment

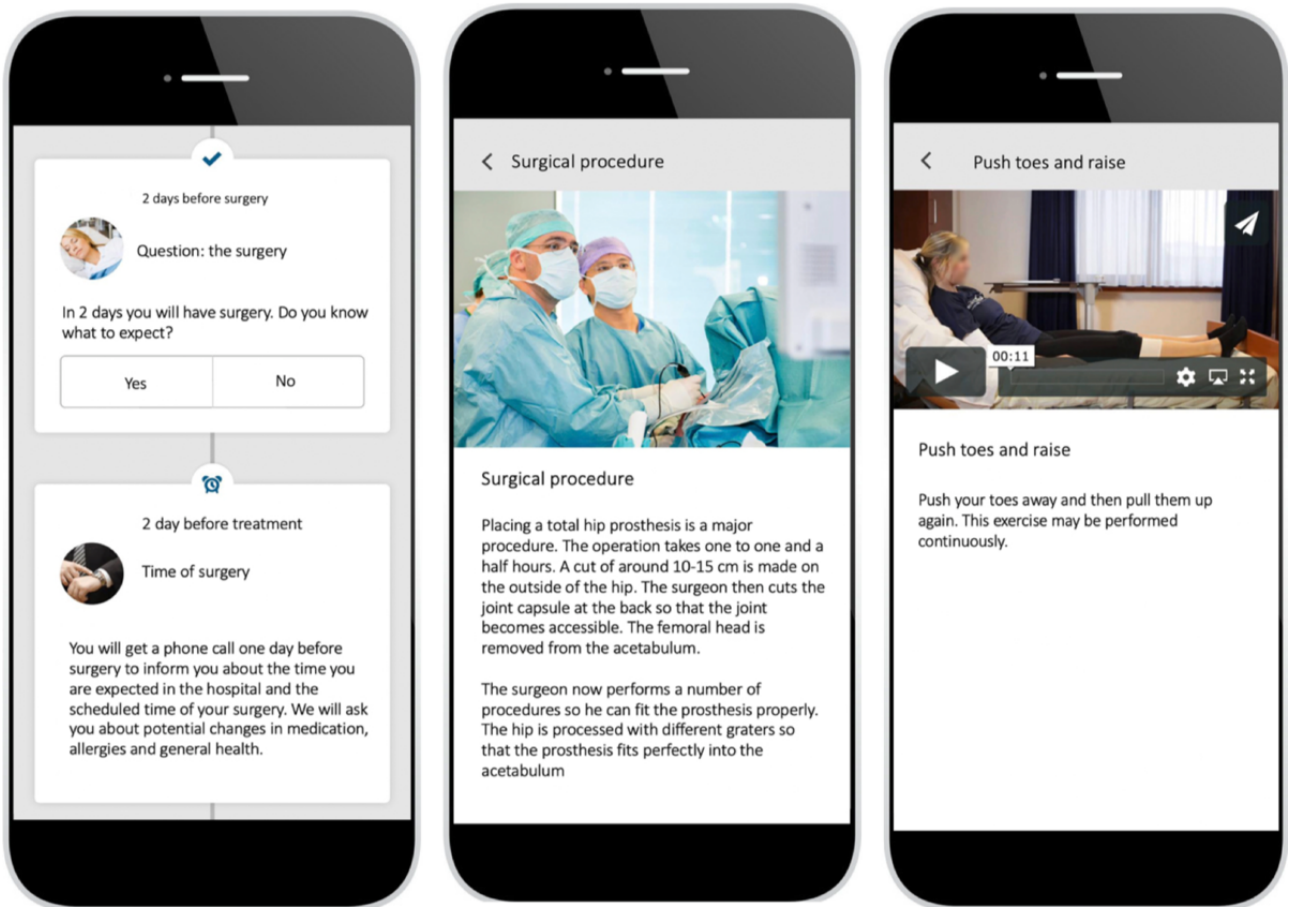
Participants were recruited in a multidisciplinary clinic (Kliniek ViaSana). Patients were eligible if they were older than 18 years and undergoing surgery for a musculoskeletal disorder. All patients were routinely informed about the app by the medical team, a brochure, and a banner in the waiting room. Patients were included if they used the Patient Journey app during their operative period and completed the web-based survey.

The Patient Journey App

The template of the Patient Journey app was developed by Interactive Studios [12]. The content was developed specifically for the various health care paths in the clinic by the medical team and can be downloaded for free on a mobile device. The app aims to provide optimal patient information and to improve adherence and self-management.

The different health care paths in the app included total hip replacement, knee replacement, anterior cruciate ligament reconstruction, knee arthroscopy, high tibial osteotomy, lumbar diskectomy, rotator cuff repair, acromioplasty, femoral osteotomy, patellar stabilization, Morton neuroma, hallux valgus/rigidus, exostosis, and talocrural arthrodesis. The app is divided into 5 categories: (1) general information about the clinic and the surgeons, (2) preoperative medical and practical information (eg, medical information and anatomy, preoperative exercises, procedures), (3) information about the stay in the hospital (eg, anesthetics, surgical intervention, exercises, advice to be active), (4) homecoming information (eg, information about possible complications, medication, sleep), and (5) information about the rehabilitation process (eg, exercises, functional instructions). App users can decide to receive push notifications. All health care paths contained specific videos with exercises and functional instructions. An example of the user app interface is presented in Figure 1.

Figure 1. Patient Journey user interface.



Data Collection

Eligible participants were invited by email. The email contained a link to the digital survey. Data were collected by MailPlus (Spotler), a program designed to manage surveys [13]. Eligible participants who did not complete the survey after 1 week received an electronic reminder. Completion of the survey took approximately 15 to 20 minutes.

Primary Outcome Measurements

The primary outcomes were (1) usability and (2) specific attitude of eHealth users toward the app. Usability was measured with the System Usability Scale (SUS) [14,15]. The SUS is a reliable and robust 10-item questionnaire and scores on a 5-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) [14,16]. The total SUS score (0 to 100) can be interpreted as not acceptable (0-64), acceptable (65 to 84), or excellent (85 to 100) [17,18]. The attitude of eHealth users toward the app was measured with part 2 of the eHealth impact questionnaire (eHIQ), which includes 3 subscales: (1) Confidence and identification (9 items), (2) Information and presentation (8 items), and (3) Understanding and motivation (9 items) [19]. The eHIQ uses a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). *Confidence and Identification* measures to what extent using the app has affected the confidence of app users in discussing and managing their health with others and whether individuals could identify with others who use the app [19]. *Information and Presentation* measures the ease of use from the user's perspective [19]. *Understanding and Motivation* measures whether respondents felt reassured, understood their condition, and felt motivated to manage their health [19]. We transformed the total scores for each subscale to a scale of 0 to 100. A score of 65 or higher was considered as a positive attitude with higher scores representing a more positive attitude toward the app [20,21]. All subscales have good internal consistency, test-retest reliability, and construct validity (Cronbach α=.88-.90) [19,20].

Secondary Outcome Measurements

The secondary outcomes were positive and negative user experiences. These were measured by overall satisfaction with the app, most appreciated and used parts of the app, satisfaction with the amount of information provided, whether the app was recommendable, reusability, supportiveness, and strengths and limitations of the app. *Satisfaction with the app* was evaluated with a numeric rating scale ranging from 0 (absolutely not satisfied) to 10 (absolutely satisfied). *The most appreciated and used parts of the app* and *the amount of information provided* were evaluated with multiple choice questions. Supportiveness, whether the app was recommendable, and reusability were measured with a 5-point Likert scale. *Supportiveness* was defined as the extent to which the respondent felt that the app was supportive in addition to the information given by health professionals and ranged from 1 (very poor) to 5 (excellent). Whether the app was *recommendable* was defined as ranging from 1 (not recommendable) to 5 (highly recommendable). *Reusability* was defined as the extent to which the respondent would use the app again if they had another surgery and ranged

from 1 (strongly disagree) to 5 (strongly agree). Strengths and limitations were gathered via open-ended questions.

Statistical Analysis

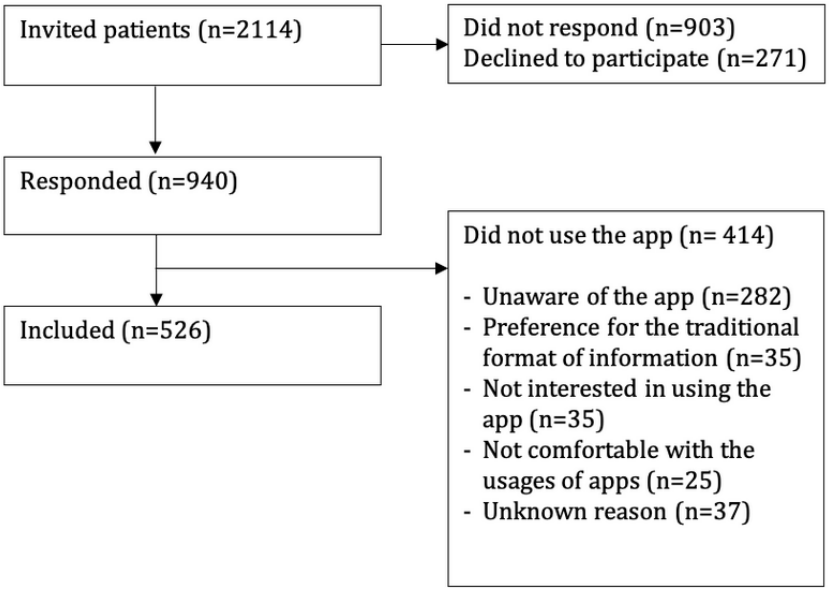
Descriptive analyses were performed to present patient characteristics and user-friendliness outcomes. Data were checked for normality using the Q-Q plots, histograms, and the Kolmogorov-Smirnov test. For the primary outcomes, significant differences between the different health care paths were tested by 1-way analysis of variance with Tukey posthoc tests (for continuous variables with a normal distribution) or the Kruskal-Wallis H test with Dunn posthoc tests (for continuous variables with a violation of normality). Posthoc Bonferroni correction was applied for multiple comparisons. For all statistical tests, α=.05 was used to determine statistical significance. All analyses were performed in SPSS (version 25.0; IBM Corporation). The positive and negative user experiences of the app were analyzed by descriptive statistics. Strengths and limitations were analyzed by using inductive thematic content analysis by 2 investigators (SJW, GGMSP) [22]. The thematic content analysis was inductive which means that no preexisting theory was imposed on the analysis. Two investigators reviewed the entire data set independently to get familiar with the responses. Subsequently, they coded the data independently and generated themes in a consensus meeting. In a second meeting (SJW, GGMSP, MWC), consensus was reached.

Results

Study Population

The survey was sent to 2114 possible participants, of whom 940 (46.7%) responded, 271 (13.4%) declined to participate, and 903 (39.9%) did not respond (Figure 2). Of the 940 participants who responded, 526 (56.0%) had used the app during their perioperative period.

Figure 2. Study flowchart.



The median age of the app users was 59.0 years (IQR 50.0-66.0), and 267 (50.8%) were female. Table 1 shows the patient characteristics of the app users and the number of participants in the different health care paths. More people who used the app were younger (P<.001), more educated (P=.01), and more frequently in paid employment (P<.001) compared to those who did not use the app.

Table 1. Patient characteristics of app users.

Variable	App users (n=526)
Gender, n (%)	
Male	259.0 (49.2)
Female	267.0 (50.8)
Age in years, median (IQR)	59.0 (50.0-66.0)
Educational level, n (%)	
Low (lower vocational education)	124.0 (23.6)
Middle (high school or secondary vocational education)	227.0 (43.2)
High (higher professional education and/or university)	175.0 (33.2)
Duration of symptoms before surgery in months, median (IQR)	22.0 (7.0-36.0)
Paid employment, n (%)	
Yes	320.0 (60.8)
No	206.0 (39.2)
Health care paths, n (%)	
Total hip replacement	89.0 (16.9)
Knee replacement	164.0 (31.2)
Anterior cruciate ligament reconstruction	56.0 (10.6)
Knee arthroscopy	47.0 (8.9)
High tibial osteotomy	23.0 (4.4)
Lumbar diskectomy	17.0 (3.2)
Rotator cuff repair	30.0 (5.7)
Acromioplasty	14.0 (2.7)
Rest group[a]	86.0 (16.3)

[a]Rest group includes shoulder arthroplasty, femoral osteotomy, patellar stabilization, Morton neuroma, hallux valgus/rigidus, exostosis, talocrural arthrodesis.

Primary Outcomes

Participants rated the app as highly usable (SUS: median 85.0, IQR 72.5-92.5; Table 2; Figure 3), and they had positive attitudes regarding information and presentation (eHIQ Information and Presentation: median 78.1, IQR 68.8-84.4; Table 3, Figure 4a). No significant differences between different health care paths were observed for usability (χ^2_8=15.5, *P*=.07).

Table 2. System Usability Scale scores (0 to 100).

Health care paths (n=526)	Usability, median (IQR)
Total hip replacement (n=89)	85.0 (71.3-95.0)
Knee replacement (n=164)	85.0 (72.5-95.0)
Anterior cruciate ligament reconstruction (n=56)	80.0 (70.6-85.0)
Knee arthroscopy (n=47)	82.5 (75.0-90.0)
High tibial osteotomy (n=23)	87.5 (77.5-95.0)
Lumbar diskectomy (n=17)	87.5 (71.3-91.3)
Rotator cuff repair (n=30)	87.5 (77.5-95.0)
Acromioplasty (n=14)	90.0 (78.8-98.1)
Rest group (n=86)[a]	85.0 (72.5-97.5)
Total group	85.0 (72.5-92.5)

[a]Rest group includes shoulder arthroplasty, femoral osteotomy, patellar stabilization, Morton neuroma, hallux valgus/rigidus, exostosis, talocrural arthrodesis.

Figure 3. Usability.

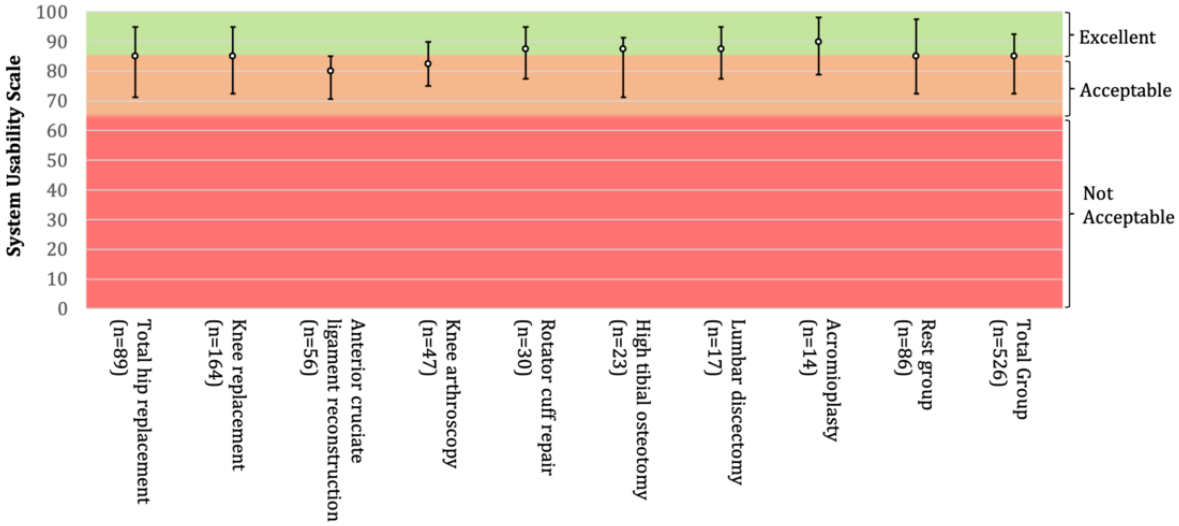
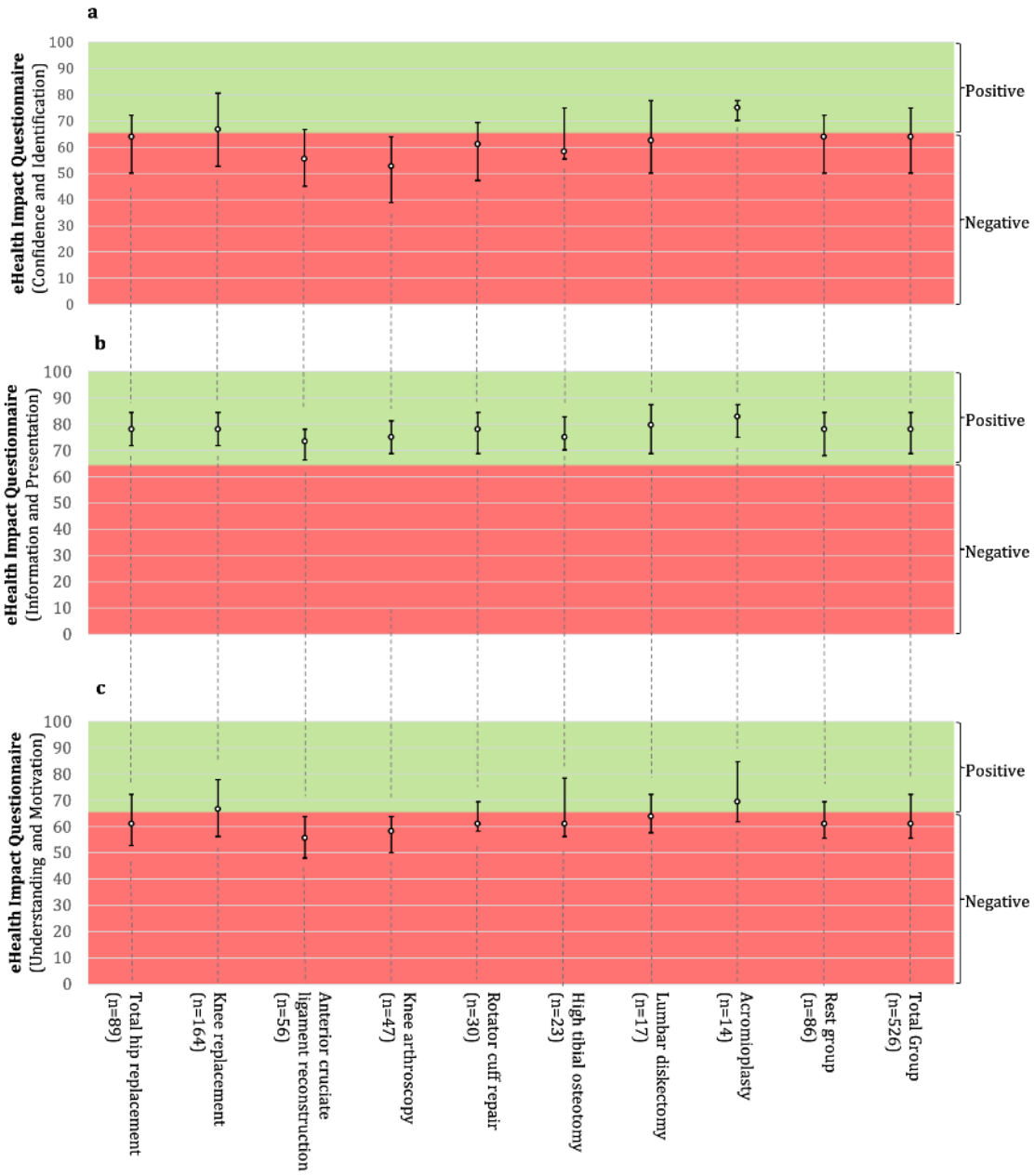


Table 3. Attitude toward the app (eHealth Impact Questionnaire, 0 to 100).

Health care paths (n=526)	Confidence and Identification, median (IQR)	Information and Presentation, median (IQR)	Understanding and Motivation, median (IQR)
Total hip replacement (n=89)	63.9 (50.0-72.2)	78.1 (71.9-84.4)	61.1 (52.8-72.2)
Knee replacement (n=164)	66.7 (52.8-80.6)	78.1 (52.8-80.6)	66.7 (56.3-77.8)
Anterior cruciate ligament reconstruction (n=56)	55.6 (45.1-66.7)	73.4 (66.4-78.1)	55.6 (47.9-63.9)
Knee arthroscopy (n=47)	52.8 (38.9-63.9)	75.0 (68.8-81.3)	58.3 (50.0-58.3)
High tibial osteotomy (n=23)	61.1 (47.2-69.4)	78.1 (68.8-84.4)	61.1 (58.3-69.4)
Lumbar diskectomy (n=17)	58.3 (55.6-75.0)	75 (70.3-82.8)	61.1 (56.3-78.5)
Rotator cuff repair (n=30)	62.5 (50.0-77.8)	79.7 (68.8-87.5)	63.9 (57.6-72.2)
Acromioplasty (n=14)	75.0 (70.1-77.8)	82.8 (75.0-87.5)	69.4 (61.8-84.7)
Rest group (n=86)[a]	63.9 (50.0-72.2)	78.1 (68.0-84.4)	61.1 (55.6-69.4)
Total group	63.9 (50.0-75.0)	78.1 (68.8-84.4)	61.1 (55.6-72.2)

[a]Rest group includes shoulder arthroplasty, femoral osteotomy, patellar stabilization, Morton neuroma, hallux valgus/rigidus, exostosis, talocrural arthrodesis.

Figure 4. Attitude toward the app: (a) Confidence and Identification, (b) Information and Presentation, and (c) Understanding and Motivation.



Participants stated that using the app did not increase their confidence in discussing and managing health with others and their feeling of identification with others (eHIQ Confidence and Identification: median 63.9, IQR 50.0-75.0) (Table 3, Figure 4a). They did not feel more reassured, did not understand their condition better, and did not feel more motivated to manage their health by using the app (eHIQ Understanding and Motivation: median 61.1, IQR 55.6-72.2; Table 3, Figure 4c). Significant differences between the various health care paths were found for the Confidence and Identification subscale (χ^2_8=44.6, P<.001), Information and Presentation (χ^2_8=17.3, P=.03), and Understanding and Motivation (χ^2_8=35.4, P<.001) subscales (Table 4). Posthoc Bonferroni comparisons showed that participants who underwent anterior cruciate ligament reconstruction scored lower than participants who underwent knee replacement (P<.001) or acromioplasty (P=.03), and similarly, participants who underwent knee arthroscopy scored lower than participants who underwent knee replacement (P<.001) or acromioplasty (P=.02) on the Confidence and Identification subscale. Participants who underwent an anterior cruciate ligament reconstruction (P<.001) or a knee arthroscopy (P=.03) scored lower than people who underwent a knee replacement on the Understanding and Motivation subscale.

Table 4. Comparison results of the health care paths.

Health care path comparison	*P* values		
	Confidence and identification	Information and presentation	Understanding and motivation
Difference between the health care paths	<.001	.03	<.001
Bonferroni posthoc analysis			
Anterior cruciate ligament reconstruction vs knee replacement	<.001	—[a]	<.001
Anterior cruciate ligament reconstruction vs acromioplasty	.03	—	—
Knee arthroscopy vs knee replacement	<.001	—	.03
Knee arthroscopy vs acromioplasty	.02	—	—

[a]Not tested because there was no difference between the health care paths.

Secondary Outcomes

App users reported a median score of 9.0 (IQR 8.0-9.0) for overall satisfaction with the app. The delivery of timely information (244/526, 46.4%) and the exercise videos (135/526, 25.7%) were the most appreciated parts of the app; 93% (475/526) would recommend the app to other patients, 86.1% (453/526) found the app supportive in addition to the information given by health professionals, and 87.3% (459/526) found the amount of information exactly enough. They appreciated the information about the stay in the hospital the least and the preoperative information the most (Multimedia Appendix 1).

The results of the inductive thematic content analyses are shown in Table 5. Important strengths related to the theme *content and information* were the clear information and instructions, timely information, and clear videos with exercises and instruction. Participant 219 wrote, “I knew exactly which exercises or activities I was able to perform each day.” Limitations belonging to this theme were that information about complications and pain medication use was lacking, an abnormal clinical course was scarcely presented, and information was not completely in line with the information provided by the medical specialist and not always up to date. Participant 293 wrote: “I found the timeline too optimistic and the information given was based on a protocol that did not fit with my situation.” Participant 27 responded, “I missed information about pain medication use.” Important strengths related to the theme *expectations and experiences* were the guidance and preparation for the surgery and rehabilitation, additional supervision and the usefulness of the app. Participant 377 responded, “The app helps you what you may expect and when.” Participant 30 wrote, “The app gave me the confidence in the journey.” Experienced limitations were that the app was not entirely personalized and missed reference data from peers. Participant 52 stated, “adding comparisons with others could provide more confidence in my personal recovery.” Participant 373 wrote, “Recovery is based on the average patient and not the individual one.” Strengths regarding the theme *technical performance* were the simplicity of downloading the app and receiving of push notifications. Patient 379 wrote, “I liked the easy way in which push notifications could be switched on and off.” Limitations were that the app sometimes jumped back and did not continue with the current phase, interaction with clinicians and access to personal electronic health records. Participant 195 stated, “It would be nice to have insight in my personal health records and the possibility to ask questions via the app.”

Table 5. Results of the inductive thematic content analyses.

Core theme and strengths	Limitations
Content and information	
• Clear information and instructions • Timely information • Useful to read back information • Clear videos with exercises and instructions	• Too optimistic information • Information about complications, pain medication use and an abnormal course are scarcely presented • Not completely in line with the information by the medical specialist • Information was not always up-to-date
Expectations and experiences	
• Optimal guidance/preparation for surgery and rehabilitation • Additional supervision • Easy to use • Clear expectations and guidelines	• Not entirely personalized • No reference data from peers
Technical performance	
• Simplicity to download the app and receive a push notification	• No interaction with clinicians • No access to personal electronic health record • App jumped back to a previous phase instead of continuing with the current phase

Discussion

Principal Results

We aimed to evaluate user-friendliness in terms of usability and attitudes of users toward the Patient Journey app. The secondary aim was to evaluate positive and negative user experiences. Indicated as the main findings, the usability of the Patient Journey app scores excellent and users have positive attitudes toward the Information and Presentation provided via the app. However, the app did not adequately improve confidence in discussing health with others and motivation to manage health. These outcomes differed between the various health care paths with lower scores in the anterior cruciate ligament reconstruction path and knee arthroscopy path. Most users would recommend the app to other patients and found the app supportive in addition to the information given by health professionals

The results of the thematic analyses provided insight into potential reasons why the Confidence and Identification and Understanding and Motivation subscale scores were below the recommended value [20,21]. Lack of personalized information, protocols based on the average patient, no interaction with clinicians, and missing reference data of peers were potential reported explanations. Previous research showed that the usage of interactive systems, videoconferencing sessions, and phone counselling favors in improving physical function, disability, and pain in comparison to conventional methods of information delivery following total knee and hip replacement [23]. Adding advanced telerehabilitation functions, such as including personal logs with appointments and a more personalized prognosis, or chat interactions with a physician or physiotherapist could probably increase a positive attitude of users toward the app.

Moreover, overly optimistic information, the scarcity of information about pain medication use, and how to act in case of a complication or deviation of the described clinical course could have led to the lower scores on the Confidence and Identification and Understanding and Motivation scales. Recent studies have shown that mHealth apps are promising tools in the guidance of pain control and opiate use and are effective in reducing pain medication intake [24,25]. It is therefore assumed that implementing pain measurements and content how to reduce pain medication into the app could reinforce a positive attitude of users toward the app.

An interesting finding is that participants who underwent an anterior cruciate ligament reconstruction and knee arthroscopy scored more negative on the Confidence and Identification and the Understanding and Motivation scale compared to other specific health care paths. Additional posthoc analyses revealed that participants in these groups were significantly younger than the other participants. Previous research also showed that middle-aged and older users pay more attention to their health issues and are more motivated to take action by using mHealth to avoid illness and stay healthy [26]. Therefore, we assume that younger patients are more confident in their capabilities, less motivated to manage their health, and less focused on specific health management.

Furthermore, following an intensive guided rehabilitation program after anterior cruciate ligament reconstruction could lead to higher levels of motivation and a better understanding of their condition. This may reduce the need for an app.

Comparison With Prior Work

Although the Patient Journey app is widely used and implemented, no previous study has assessed its user-friendliness. Other research described the user-friendliness of various types of mHealth interventions having dissimilar purposes in different health care settings [27-32]. These studies [27-32] also demonstrated that mHealth apps are highly feasible and acceptable to users. No previous studies assessed the user-friendliness of mHealth tools for the perioperative period for musculoskeletal surgery. A recent systematic review [10] evaluated patients' experiences on the use of perioperative mHealth apps; these authors found that mHealth can serve as an important tool for patient engagement in education about

their condition and procedure. Moreover, mHealth apps can reduce inconsistencies between information given by health care providers [10]. Although the information provided and instructions were one of the strengths of the Patient Journey App, our qualitative analysis showed that the information provided was not always in line with that provided by the medical specialist. Comparable with our findings, reported weaknesses for perioperative mHealth use were patients' lack of confidence, lack of personalized information, and often overly optimistic information which could lead to an overestimation of the patients' course [10]. The timely information as provided by the Patient Journey app helps people to comprehend information and has positive effects on the patients' levels of knowledge, satisfaction, clinical outcomes, and health care economics [9].

A general strength and important motivator for mHealth users is the accessibility of specific information that could increase knowledge about their condition [31,33]. Nevertheless, an important concern regarding trustworthiness is that this information is not always up-to-date and valid [31]. Other important factors in line with those in previous research are the lack of personalization, peer support, and integration of functionalities that enhance the interaction with clinicians [30,31]. To increase the relevance of app use, it is preferable that mHealth apps include diverse functions that enable patients to personalize and tailor them to meet their needs [31,32]. Furthermore, peer support can enhance patient socialization by providing social support, and facilitating 2-way communication with clinicians could increase patient engagement and therefore seems to be a great promise of mHealth [31,34]. In contrast to our findings, mHealth apps for patients with chronic diseases can increase feelings of managing health-related behavior by making users feel more reassured and empowered [27,31]. Most of our participants, however, did not feel more confident in managing their health by using the Patient Journey App. Potential differences could be explained by the type of participants (people with chronic diseases versus people with musculoskeletal disorders scheduled for surgery) and engagement in self-management (people who undergo musculoskeletal surgery may have less need to be engaged in self-management, especially during the stay in the hospital compared to patients with chronic health issues) [27,31]. Patients who are highly engaged in self-management experience the use of mHealth apps as more beneficial than others [31].

Limitations

This study has several limitations. First, the number of participants in the different health care paths varied, and this could have led to imprecise results in health care paths with small sample sizes. Second, we used inductive thematic content analyses based on open-ended questions for the secondary outcomes. Semistructured interviews could have helped to define areas that could be further explored and would have given more detailed information about some themes [35]. The representativeness of the study might be biased as participants who used the app were statistically significantly younger ($P<.001$), higher educated ($P=.01$), and had more paid jobs ($P<.001$) compared to those who did not use the app. Moreover, most of our participants belonged to the middle-age group. It is unclear whether the results would have been different in younger or older age groups as different age groups may have different experiences of app usability and different expectations for how apps should function [26].

Despite these limitations, we believe that this study does provide novel insights into the user-friendliness of the mHealth app in the perioperative musculoskeletal period and that the results are of clinical importance for app users, clinicians, mHealth app developers, and researchers.

Conclusion

The Patient Journey app is a usable, highly informative, and presentable tool to inform patients with a musculoskeletal disorder during their perioperative period. For participants in most health care paths, using the app did not improve their confidence in discussing their health or reassurance in managing their health. However, the development of utilities that can offer reference data from peers, interaction with clinicians, and more insight into pain could further increase the user-friendliness of the app.

Acknowledgments

The authors would like to thank Mike Gabriël of Kliniek ViaSana, for designing the survey, and all patients who participated in the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Secondary outcomes.
[DOCX File , 23 KB - humanfactors_v8i1e20694_app1.docx]

References

1. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. J Med Internet Res 2014 Apr 16;16(4):e110 [FREE Full text] [doi: 10.2196/jmir.2790] [Medline: 24739471]

2. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of web-based vs. non-web-based interventions: a meta-analysis of behavioral change outcomes. J Med Internet Res 2004 Nov 10;6(4):e40 [FREE Full text] [doi: 10.2196/jmir.6.4.e40] [Medline: 15631964]
3. Walker C, Swerissen H, Belfrage J. Self-management: its place in the management of chronic illnesses. Aust Health Rev 2003;26(2):34-42. [doi: 10.1071/ah030034a] [Medline: 15368834]
4. Sandberg E, Sharma R, Sandberg W. Deficits in retention for verbally presented medical information. Anesthesiology 2012 Oct;117(4):772-779 [FREE Full text] [doi: 10.1097/ALN.0b013e31826a4b02] [Medline: 22902965]
5. Timmers T, Janssen L, Pronk Y, van der Zwaard BC, Koëter S, van Oostveen D, et al. Assessing the efficacy of an educational smartphone or tablet app with subdivided and interactive content to increase patients' medical knowledge: randomized controlled trial. JMIR mHealth uHealth 2018 Dec 21;6(12):e10742 [FREE Full text] [doi: 10.2196/10742] [Medline: 30578185]
6. Khaleel I, Wimmer BC, Peterson GM, Zaidi STR, Roehrer E, Cummings E, et al. Health information overload among health consumers: a scoping review. Patient Educ Couns 2020 Jan;103(1):15-32. [doi: 10.1016/j.pec.2019.08.008] [Medline: 31451363]
7. Houts PS, Doak CC, Doak LG, Loscalzo MJ. The role of pictures in improving health communication: a review of research on attention, comprehension, recall, and adherence. Patient Educ Couns 2006 May;61(2):173-190. [doi: 10.1016/j.pec.2005.05.004] [Medline: 16122896]
8. Bol N, Smets EM, Rutgers MM, Burgers JA, de Haes HC, Loos EF, et al. Do videos improve website satisfaction and recall of online cancer-related information in older lung cancer patients? Patient Educ Couns 2013 Sep;92(3):404-412. [doi: 10.1016/j.pec.2013.06.004] [Medline: 23820196]
9. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using smartphone and tablet apps: systematic review. J Med Internet Res 2020 Apr 13;22(4):e17342 [FREE Full text] [doi: 10.2196/17342] [Medline: 32281936]
10. De La Cruz Monroy MFI, Mosahebi A. The use of smartphone applications (apps) for enhancing communication with surgical patients: a systematic review of the literature. Surg Innov 2019 Apr;26(2):244-259. [doi: 10.1177/1553350618819517] [Medline: 30602332]
11. Simpao AF, Lingappan AM, Ahumada LM, Rehman MA, Gálvez JA. Perioperative smartphone apps and devices for patient-centered care. J Med Syst 2015 Sep;39(9):102. [doi: 10.1007/s10916-015-0310-7] [Medline: 26265239]
12. Patient Journey App. Website in Dutch. URL: https://patientjourneyapp.com/nl [accessed 2018-11-12] [WebCite Cache ID 74aGhHznBto]
13. Spotler. Webpage in Dutch. URL: https://spotler.com/?gclid=Cj0KCQiA3b3gBRDAARIsAL6D-N-CGhIzj1h2q24QAR-xrSklTs7S3xRhm0WQWy7qPjgyVHlcNhw14s4aAmVbEALw_wcB [accessed 2018-12-18] [WebCite Cache ID 74aOrNSMa]
14. Brooke J. SUS: a 'quick and dirty' usability scale. In: Usability Evaluation In Industry. London: CRC Press; 1996.
15. Lewis JS. The factor structure of the system usability scale. In: Kurosu M, editor. Human Centered Design. Berlin, Germany: Springer; May 27, 2009.
16. Dicianno BE, Parmanto B, Fairman AD, Crytzer TM, Yu DX, Pramana G, et al. Perspectives on the evolution of mobile (mHealth) technologies and application to rehabilitation. Phys Ther 2015 Mar;95(3):397-405 [FREE Full text] [doi: 10.2522/ptj.20130534] [Medline: 24925075]
17. Daemen E, Flinsenberg I, Van LE, Cuppen R, Rajae-Joordens R. The effect of experience on system usability scale ratings. J Usability Stud 2012;7(Issue 2):56-67.
18. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, California: Createspace Independent Publishing Platform; Apr 20, 2011.
19. Kelly L, Ziebland S, Jenkinson C. Measuring the effects of online health information: Scale validation for the e-Health Impact Questionnaire. Patient Educ Couns 2015 Nov;98(11):1418-1424 [FREE Full text] [doi: 10.1016/j.pec.2015.06.008] [Medline: 26162953]
20. Neijenhuijs KI, van der Hout A, Veldhuijzen E, Scholten-Peeters GGM, van Uden-Kraan CF, Cuijpers P, et al. Translation of the eHealth Impact Questionnaire for a population of Dutch electronic health users: validation study. J Med Internet Res 2019 Aug 26;21(8):e13408 [FREE Full text] [doi: 10.2196/13408] [Medline: 31452516]
21. Talboom-Kamp E, Tossaint-Schoenmakers R, Goedhart A, Versluis A, Kasteleyn M. Patients' attitudes toward an online patient portal for communicating laboratory test results: real-world study using the eHealth Impact Questionnaire. JMIR Form Res 2020 Mar 04;4(3):e17060 [FREE Full text] [doi: 10.2196/17060] [Medline: 32024632]
22. Green J, Thorogood N. Qualitative Methods for Health Research. London: Sage Publications; 2004.
23. Pastora-Bernal JM, Martín-Valero R, Barón-López FJ, Estebanez-Pérez MJ. Evidence of benefit of telerehabitation after orthopedic surgery: a systematic review. J Med Internet Res 2017 Apr 28;19(4):e142 [FREE Full text] [doi: 10.2196/jmir.6836] [Medline: 28455277]
24. Pronk Y, Peters MCWM, Sheombar A, Brinkman J. Effectiveness of a mobile ehealth app in guiding patients in pain control and opiate use after total knee replacement: randomized controlled trial. JMIR Mhealth Uhealth 2020 Mar 13;8(3):e16415 [FREE Full text] [doi: 10.2196/16415] [Medline: 32167483]

25. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
26. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? a meta-analysis and the moderating role of age. International Journal of Information Management 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
27. Seto E, Morita P, Tomkun J, Lee TM, Ross H, Reid-Haughian C, et al. Implementation of a heart failure telemonitoring system in home care nursing: feasibility study. JMIR Med Inform 2019 Jul 26;7(3):e11722 [FREE Full text] [doi: 10.2196/11722] [Medline: 31350841]
28. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
29. Jeffrey B, Bagala M, Creighton A, Leavey T, Nicholls S, Wood C, et al. Mobile phone applications and their use in the self-management of type 2 diabetes mellitus: a qualitative study among app users and non-app users. Diabetol Metab Syndr 2019;11:84 [FREE Full text] [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
30. Juarascio A, Goldstein S, Manasse S, Forman E, Butryn M. Perceptions of the feasibility and acceptability of a smartphone application for the treatment of binge eating disorders: qualitative feedback from a user population and clinicians. Int J Med Inform 2015 Oct;84(10):808-816 [FREE Full text] [doi: 10.1016/j.ijmedinf.2015.06.004] [Medline: 26113461]
31. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mhealth apps: meta-ethnographic review of qualitative studies. JMIR mHealth uHealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
32. Bauer A, Iles-Shih M, Ghomi R, Rue T, Grover T, Kincler N, et al. Acceptability of mhealth augmentation of collaborative care: a mixed methods pilot study. Gen Hosp Psychiatry 2018;51:22-29 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.11.010] [Medline: 29272712]
33. Peng W, Kanthawala S, Yuan S, Hussain S. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
34. Martinez P. A qualitative study on patient perceptions towards mHealth technology among high risk, chronic disease patients. Harvard Medical School. 2015. URL: http://nrs.harvard.edu/urn-3:HUL.InstRepos:17295915 [accessed 2021-01-14]
35. Gill P, Stewart K, Treasure E, Chadwick B. Methods of data collection in qualitative research: interviews and focus groups. Br Dent J 2008 Mar 22;204(6):291-295. [doi: 10.1038/bdj.2008.192] [Medline: 18356873]

Abbreviations

eHIQ: eHealth Impact Questionnaire
mHealth: mobile health
SUS: System Usability Scale

Edited by A Kushniruk; submitted 26.05.20; peer-reviewed by I Contreras, A Budenz; comments to author 17.07.20; revised version received 06.10.20; accepted 23.12.20; published 08.02.21.

Please cite as:
Willems SJ, Coppieters MW, Pronk Y, Diks MJF, van der Heijden KWAP, Rooker S, Scholten-Peeters GGM
A Clinical Journey Mobile Health App for Perioperative Patients: Cross-sectional Study
JMIR Hum Factors 2021;8(1):e20694
URL: http://humanfactors.jmir.org/2021/1/e20694/
doi:10.2196/20694
PMID:33555262

Original Paper

Perspectives of Trial Staff on the Barriers to Recruitment in a Digital Intervention for Psychosis and How to Work Around Them: Qualitative Study Within a Trial

Stephanie Allan[1], MA; Hamish Mcleod[1], PhD; Simon Bradstreet[1], PhD; Imogen Bell[2], PhD; Helen Whitehill[1], BA; Alison Wilson-Kay[1], MSc; Andrea Clark[1], MSc; Claire Matrunola[1], MSc; Emma Morton[3], PhD; John Farhall[4], PhD; John Gleeson[5], PhD; Andrew Gumley[1], PhD

[1]University of Glasgow, Glasgow, United Kingdom
[2]Orygen, Melbourne, Australia
[3]University of British Columbia, Vancouver, BC, Canada
[4]La Trobe, Melbourne, Australia
[5]Australian Catholic University, Melbourne, Australia

Corresponding Author:
Stephanie Allan, MA
University of Glasgow
Glasgow Mental Health Research Facility
Glasgow, G20 0XA
United Kingdom
Phone: 44 01412113939
Email: stephanie.allan@glasgow.ac.uk

Abstract

Background: Recruitment processes for clinical trials of digital interventions for psychosis are seldom described in detail in the literature. Although trial staff have expertise in describing barriers to and facilitators of recruitment, a specific focus on understanding recruitment from the point of view of trial staff is rare, and because trial staff are responsible for meeting recruitment targets, a lack of research on their point of view is a key limitation.

Objective: The primary aim of this study was to understand recruitment from the point of view of trial staff and discover what they consider important.

Methods: We applied pluralistic ethnographic methods, including analysis of trial documents, observation, and focus groups, and explored the recruitment processes of the EMPOWER (Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-being, Engagement, and Recovery) feasibility trial, which is a digital app–based intervention for people diagnosed with schizophrenia.

Results: Recruitment barriers were categorized into 2 main themes: service characteristics (lack of time available for mental health staff to support recruitment, staff turnover, patient turnover [within Australia only], management styles of community mental health teams, and physical environment) and clinician expectations (filtering effects and resistance to research participation). Trial staff negotiated these barriers through strategies such as emotional labor (trial staff managing feelings and expressions to successfully recruit participants) and trying to build relationships with clinical staff working within community mental health teams.

Conclusions: Researchers in clinical trials for digital psychosis interventions face numerous recruitment barriers and do their best to work flexibly and to negotiate these barriers and meet recruitment targets. The recruitment process appeared to be enhanced by trial staff supporting each other throughout the recruitment stage of the trial.

(JMIR Hum Factors 2021;8(1):e24055) doi:10.2196/24055

KEYWORDS

recruitment; schizophrenia; mHealth; psychosis; mental health

Introduction

Background

To better understand how interventions could be developed, evaluated, and implemented in routine care, it is important to fully understand which aspects of the implementation of randomized control trials (RCTs) are most challenging [1]. All RCTs must recruit participants for interventions to be tested [2]. However, recruitment into RCTs can be very difficult and is possibly the biggest challenge within clinical research [3], with many RCTs failing to reach their recruitment targets [4]. Delayed recruitment can lead to additional costs [5], and underpowered clinical trials can threaten the empirical value of intervention research [6]. Systematic reviews of recruitment barriers have helped uncover specific barriers for recruiting ethnic minority populations [7], within HIV trials [8] and cancer trials [9]. However, reviews are only possible if primary data are collected and shared. Digital interventions are becoming popular for increasing access to treatments; however, little is known about the nature of specific recruitment barriers in these trials [10]. Beyond widespread societal concern about the negative impacts of digital technology within daily life [11], there may be recruitment challenges in mental health care research, such as concerns that patients may struggle to use a digital device [12]. However, systematic review evidence suggests that these effects are not yet understood because trial recruitment is not covered in depth in studies of implementation barriers for digital interventions for psychosis [13].

Trial staff responsible for recruiting participants must implement something novel (in this case, the recruitment process for a new intervention) within a health care system that comes with existing norms, knowledge, and social practices. Trial recruitment involves interacting with diverse groups [14] including patients, clinical staff, clinical leaders, and other members of the trial team. The health care system can be described as a context in which the recruitment process must fit. Process evaluations use qualitative research to develop an understanding of how trial processes such as recruitment were delivered and received by participants and trial staff [15,16]. Context in process evaluation terms is defined as factors external to an intervention that influence clinical trial processes' delivery [17], such as recruitment. Therefore, understanding the context of recruitment is important for understanding what factors may act as barriers and facilitators in enrolling participants within a clinical trial.

Use of and interest in digital interventions is high in people diagnosed with schizophrenia [18], and digital interventions for psychosis are growing in popularity [19,20]. Currently, the ongoing COVID-19 pandemic has seen a surge in interest in using digital technologies to support people with mental health problems [21]. However, the willingness of patients to be recruited into digital intervention clinical trials is poorly understood [22,23]. People diagnosed with schizophrenia are described as a difficult-to-recruit population, more generally within clinical trials [24]. Recruitment for service users diagnosed with schizophrenia often involves approaching patients via staff; therefore, it seems particularly important to consider the role of staff within study recruitment. For example, a recent study reported that 1 in 5 mental health staff report having never recruited a service user into a research study [25].

Within trials of digital interventions, it is recommended that the recruitment of end users should be described in sufficient detail to enable readers who wish to contextualize or replicate the work [26]. Feasibility studies help establish important parameters such as the willingness of clinicians to recruit patients and the willingness of participants to be randomized [27]. Despite the importance of recruitment, CONSORT (Consolidated Standards of Reporting Trials) statements [28] do not require RCT reporting to describe recruitment in detail beyond documentation of participant flow [29,30]. The proposed CONSORT extensions [31] recommended that qualitative data be collected so that context can be more fully understood and so that future researchers may recognize relevant contextual elements (such as settings and stakeholder participation) that are necessary for the replication of findings observed within a particular trial. Reporting a more detailed examination of recruitment processes, particularly recruitment barriers [32], is suggested to be useful in interpreting trial results and developing strategies for improvement [33]. Moreover, failure to report recruitment experiences risks significant loss of a key source of knowledge. In addition, it is important to note that detailed reporting of recruitment into digital intervention studies using mobile apps is scarce [34].

Trial staff are responsible for meeting recruitment targets, which requires interacting with potential participants. This places them in a unique position to comment on the overall recruitment process and provides a narrative on (1) what happened during trial recruitment and (2) to enable researchers to make informed comment on why. Identifying barriers to recruitment has been identified as a strength of qualitative research within clinical trials [35,36]. Furthermore, qualitative research could also describe what strategies trial staff use to negotiate around recruitment barriers. However, to the best of our knowledge, there is little empirical exploration of the trial recruitment process directly from the point of view of trial staff.

Study Aims

This qualitative study within a trial (SWAT) [37] aimed to gather and analyze data to more fully understand barriers and facilitators encountered by trial staff during the recruitment process for the EMPOWER (Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-being, Engagement, and Recovery) study (described in more detail later) and to facilitate learning ahead of a full trial. Previous qualitative work conducted with carers, mental health staff, and service users suggested that recruitment barriers were hypothesized within the EMPOWER trial [12], such as service users feeling paranoid in response to digital technology and a lack of staff time to support the recruitment process. Therefore, this study aims to explore recruitment issues in some depth but was not limited to the a priori issues identified in our previous research.

EMPOWER [38] (ISRCTN: 99559262) aimed to develop and evaluate a mobile app for use with adults who experience psychosis. The EMPOWER app is a digital self-management tool (augmented with peer support) to enhance the identification

of and communication about early warning signs of relapse in people diagnosed with schizophrenia. The app enables routine self-monitoring for a variety of different experiences, including psychosis (eg, hearing voices and suspicious thoughts), anxiety, mood, self-esteem, and interpersonal support. EMPOWER participants used the app for an initial 28-day baseline period to identify their typical variation in personal well-being. Significant changes from baseline are then triaged by a clinician, and, if necessary, mental health staff are notified. EMPOWER was tested in a cluster randomized control trial (cRCT). As EMPOWER was trying to enhance communication and shared decision making between multiple stakeholders, mental health staff, service users, and carers (if relevant) were all potential participants. The feasibility of the EMPOWER intervention and study procedures was tested in a multisite trial in both Australia and the United Kingdom. The initial recruitment target was 120 service user participants (and any linked carers) and 40 mental health staff from 8 community mental health services (CMHS) before randomization of the clusters (services). During the course of the study, 8 CMHS were recruited and randomized; however, a revised recruitment target of n=86 was agreed upon and met.

In cluster trials, outcomes are usually measured at the level of the individual; however, trial procedures (such as recruitment) are applied by the research team at the level of the cluster (in this case, adult community mental health teams) [39]. When recruitment for EMPOWER began, research assistants within EMPOWER electronically screened medical records of local CMHS for potentially eligible participants and then approached key workers employed within adult community mental health teams (the cluster) who had potentially eligible participants on their case load. Therefore, developing an understanding of recruitment both within and across sites appears important in contextualizing the recruitment process in a cRCT such as EMPOWER. Full details of the intervention are reported in the protocol [38]. In a feasibility study such as EMPOWER, process evaluators are usually interested in facilitators and barriers to implementation so that strategies to enhance implementation of key processes such as recruitment can be put in place for a definitive trial [17].

Methods

Theoretical Framework

In line with the EMPOWER process evaluation protocol [40], the theoretical framework for this study was constructivism [15], which posits that knowledge is created through social interactions. The processes that occur during intervention implementation need to be understood in ways that are responsive to the complexities and intricacies of programs, people, and places [41]. Recruitment in clinical trials is a complex social action; therefore, there is unlikely to be one definitive methodology (qualitative or otherwise) that can allow us to theorize recruitment in sufficient depth [42].

The primary focus of the analysis was on achieving the a priori study aims (understanding the context of recruitment during the feasibility trial stage to refine recruitment in a full trial). Particular attention was paid to the reporting of barriers and facilitators to recruitment because this helps understand the context of recruitment. We now describe the 2 methods of the study in line with the key aim.

Ethnography

Ethnography refers to both the process and outcome of research that produces rich descriptions and interpretations of a social system from the point of view of its key social actors, including their behaviors, roles, and methods of interaction [43]. Ethnography is useful for theorizing implementation processes such as recruitment because ethnographic narratives pay attention to interconnectedness while building a holistic understanding of how systems come together as a whole [44,45]. Furthermore, ethnography is useful for developing internally valid theory by focusing on describing how people behave in the real-world context of clinical trial recruitment. Taking an ethnographic stance is advantageous in process evaluation research because it can help develop the implementation theory of key trial processes with good internal validity [46].

SA was based within the main office of EMPOWER for the full duration of trial recruitment and was able to observe trial staff both within meetings and within their daily office-based tasks during the recruitment process. Although ethnography commonly involves a researcher directly observing social processes, the examination of administrative data and study documents is important within process evaluation research [47]. Therefore, the minutes of team meetings were seen as sites for ethnographic inquiry beyond what SA recorded from observation. This was considered to be particularly useful because SA could not directly observe recruitment processes that occurred outside of the office.

Trial Staff Focus Groups

To triangulate findings from the observation-based ethnography, focus groups were held with members of trial staff who were involved in the recruitment process. The use of qualitative methods [48] and, in particular, focus groups within an RCT facilitates the understanding of the recruitment process [49]. Exploring recruitment from the point of view of the trial staff who worked on the trial and who experienced the recruitment process directly is noted to be useful because it provides insight into the reasons behind what can be observed [35]. Ethics approval for this study was received from the West of Scotland Research Ethics Service (GN16MH271 Ref: 16/WS/0225) and Melbourne Health (HREC/17/MH/97 Ref: 2017.010).

Procedure

Ethnography

SA (who was based in the UK office for the EMPOWER study) was present at the majority of weekly team meetings in the United Kingdom that were held during the recruitment process and had access to the minutes of meetings from this time. All members of the EMPOWER team who were based in Glasgow attended these meetings, with the focus of discussion being on general trial business. Recruitment procedures for both the United Kingdom and Australia were discussed in these meetings. Beyond formal meetings, SA was able to observe the work of the trial staff within the office and was privy to their discussions

and reflections on the matter for the duration of trial recruitment. SA recorded reflective notes during the recruitment process from ethnographic observations at both formal meetings and more informal *daily work* and then consolidated these into reflective memos once the recruitment period was over. SA revisited meeting minutes (n=50) for the period from August 03, 2017, when recruitment started, to July 05, 2018, when the recruitment target was achieved (n=86), to refresh their memory and wrote reflective ethnographic memos. Relevant ethnographic reflections are reported in addition to analyses from the focus groups. Observational data from meeting recordings and field notes were anonymized.

Trial Staff Focus Groups

Both focus groups were facilitated by SA (independent of the research team). One focus group was facilitated in person in Glasgow, United Kingdom, and another was facilitated remotely with the Australian team in Melbourne, who participated remotely via a secure telephone interface. Verbal informed consent was obtained before the start of each focus group. Each focus group followed a schedule of questions designed to explore barriers and facilitators to recruitment in some depth. A semistructured interview schedule was developed for broad exploration of the recruitment process from the perspective of trial staff (schedule available in prepublished protocol [40]). Both focus groups were audio recorded and then transcribed verbatim. Focus groups lasted for an hour. All focus groups were held during the typical working day for trial staff, and participation was voluntary. Data have been anonymized to protect confidentiality; all participants are simply referred to as *Participant*, with numbers being used for clarity when a textual extract has data from more than one participant.

All participants in this SWAT (through observation or focus group participation or both) were employed in the EMPOWER trial and were involved in trial recruitment (either directly or indirectly). EMPOWER was a feasibility study; therefore, the numbers reflect the relatively small pool of trial staff, which is highlighted in Table 1. NVivo [50] software was used for all analyses.

Table 1. Description of participants' characteristics.

Location	Focus group attendees	Roles
United Kingdom	6 (out of a possible 7)	Researcher, Chief Investigator, and Trial Manager
Australia	3 (out of a possible 5)	Principal Investigator, Researchers, and Trial Manager

Reflexivity

SA is a PhD student working on a process evaluation for the EMPOWER cRCT [38]. The PhD funding SA receives is independent of any funding associated with the trial. Following observations of trial staff during the recruitment process, it seemed as though the recruitment process was a key site of inquiry to more fully understand full trial feasibility. Therefore, a decision was made to undertake a small qualitative SWAT. Supervision and finalization of the coding process was done in conjunction with HM and AG, who are academic clinical psychologists, academic supervisors to SA, and investigators on the EMPOWER trial.

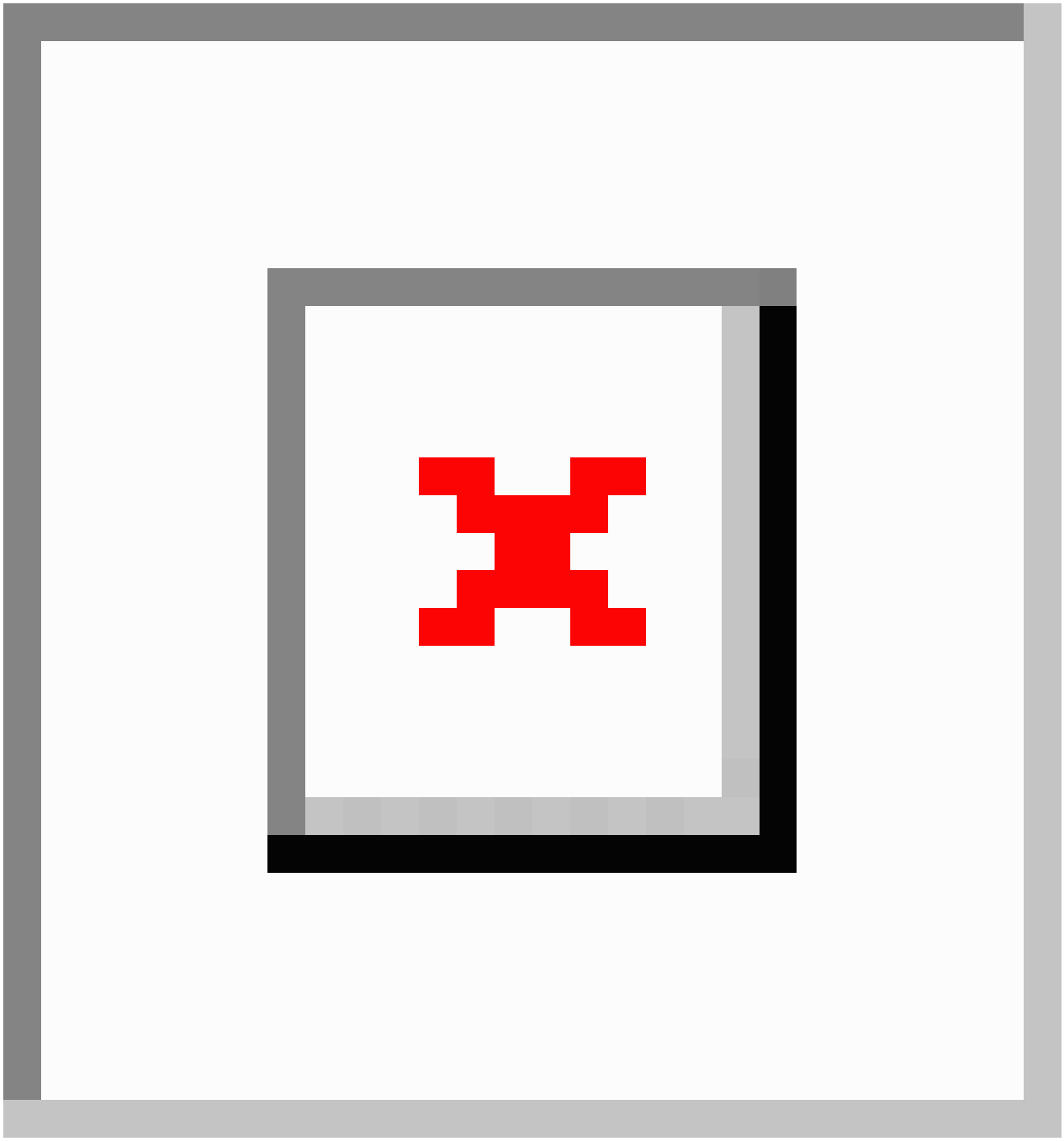
Analysis

All data, including ethnographic observations and focus group transcripts, were analyzed thematically by SA using thematic analysis, a qualitative method used to identify, analyze, and report patterns constructed within text data [51]. The first stage comprised line-by-line coding (descriptive) moving onto the second stage of coding, where descriptive codes were thematically linked together into a final set of themes. Constructivist qualitative research assumes that themes do not emerge from data but are constructed as part of a reflexive analytic process [52]. Therefore, themes will be reported as being constructed. Trial staff provided critical feedback on the rigor and validity of the thematic analysis, similar to member checking [53].

Results

Following thematic analyses of ethnographic observations and focus groups, it seemed that there were several key recruitment barriers encountered by the research team during recruitment to the trial. Beyond simply listing recruitment issues, trial staff discussed how these issues were addressed and what work was done to best negotiate these issues. To frame these discussions as distinct from merely reporting key issues, the concept of *trial work* [54] was used within a qualitative framework analysis [55]. Trial work is a broad concept related to the work done to overcome barriers during the recruitment process engagement, *buy in* to the trial across a range of stakeholders, and work involved in managing the organizational complexity necessary to reach recruitment targets [54]. Trial work appeared to be highly relevant to the aims of this study in terms of maximizing learning and understanding from the EMPOWER recruitment process. The reporting will highlight the key *recruitment barriers* and then the *trial work* used to facilitate recruitment. We summarize the themes in Figure 1 and then describe the themes and provide portions of raw data to make the analysis more transparent.

Figure 1. Thematic map of recruitment themes.



Key Recruitment Barriers

The key barriers described by trial staff into trial recruitment broadly fell into 2 main themes: service characteristics (lack of time available to mental health staff to support recruitment, staff turnover, patient turnover [within Australia only], management styles of community mental health teams, and physical environment) and clinician expectations (filtering effect and resistance to research participation).

Service Characteristics

Lack of Time Available to Mental Health Staff to Support Recruitment

Research trial staff frequently spoke about mental health staff not having much time to engage in the recruitment process. The research team was highly aware of the broader social context of low staff capacity in the face of high numbers of patient referrals in routine care with limited staff to meet demand. Trial staff at both sites made empathetic references to being aware of mental health staff working within a context of immense pressure with a lack of resources and support. During the analysis by SA, it was constructed that the trial staff in EMPOWER felt it was inevitable that structural barriers that

lead to mental health staff not having much spare time would inevitably be a barrier to trial recruitment:

> *I don't think you can relate how busy they are. And much pressure they're under. Some of the numbers we heard about in terms of new referrals into teams were quite staggering.* [Participant 1]

> *Forty. Forty referrals a week, yeah. And there doesn't seem to be any sort of throughput to accommodate that additional pressure being moved around.* [Participant 2, United Kingdom]

High Mental Health Staff Turnover

Closely linked to a lack of staff time was high staff turnover, which appeared to be systemic across both trial sites. Meeting notes and focus group data from both the United Kingdom and Australia indicated that high clinical staff turnover was a challenge to recruitment. Practically, this led to issues such as new clinical staff not being aware of the study because they were not employed when staff teams were initially told about it. Clinical staff changing jobs or taking leaves as they were unwell also appeared to be systemic issues within mental health services and was a macrolevel recruitment challenge. In the following example, a member of the EMPOWER team reflects on the impact of high staff turnover:

> *What we're seeing is the key workers [mental health staff] are very fluid, there's loads of movement, there's massive changes as to who your key worker is, there's lots of staff turnover.* [Participant, United Kingdom]

High Patient Turnover

A related subtheme (which was exclusive to Australia) was patient turnover because patients are discharged back to general practice (as evidenced in the quote below where participant alludes to "it's not only a high turnover of consumers [patients]") following the end of an acute episode of psychosis, unlike in the United Kingdom where clinical support is generally more long term for people diagnosed with schizophrenia. This was a particular barrier to recruitment because if patients were no longer in the service, they simply could not be recruited. However, this issue intersected with high clinical staff turnover, resulting in a complex barrier to recruitment into the study because the high clinical staff turnover within mental health services blocked the ability of trial staff to build relationships with clinical staff to build trust in the team and the project:

> *I think it's also worth noting that in public mental health services it's not only a high turnover of consumers [patients] but there's also a pretty high turnover of staff in some places, so you would have some clinicians that hadn't heard of it or you know were quite new around that time and that kind of translates to recruiting consumers as well in terms of the discharges and the change in people being part of the service.* [Participant, Australia]

Differences in Management Styles Within Clinical Teams

In both the United Kingdom and Australia, there were discussions about differences in management styles between the different mental health teams. In the first example, a trial team member explicitly stated that although participant numbers between sites may not have appeared too different, this obscured the challenges of having to adapt to different leadership styles across mental health teams. This was viewed as a key determinant of recruitment success:

> *I think at the big picture level the rate of recruitment wasn't particularly different and you know, [other named research assistants] might be able to say a bit more about the style of how it happens etc., there are certainly very different personality styles of managers so in terms of us managing the managers, we had to take into account that there are very different people who had a very different styles.* [Participant, Australia]

However, as pointed out in the UK site, it was not always the case that managers were those who were *pulling the strings* in terms of creating barriers to recruitment:

> *Leadership's hugely important in this. And always underestimated how much influence it has in any field, but this one no less. That the messages and the values and the attitudes that are being shared by the person who's pulling the strings is really, really important. And that person who's pulling the strings isn't necessarily always the person who is supposed to be pulling the strings.* [Participant, United Kingdom]

Differences in Physical Environment

A further important recruitment challenge stemmed from the layout of the physical premises of mental health services themselves. Although this may be unique to a particular center, the impact upon recruitment was considered by trial staff to be large. For example, 2 researchers recalled the impact of the physical layout of premises, which hindered their ability to develop relationships with staff and acted as a significant block to successful social interactions:

> *The physical environment's really problematic there [named recruitment site] as well, because they're all in small, separate offices, so it doesn't really feel like a team. So individual and...* [Participant 1]

> *There's nowhere to circulate and to talk to the nurses.* [Participant 2]

> *There's nowhere to chat amongst yourself, just to build the rapport with nurses. It was like, everyone's all huddled away in separate offices.* [Participant 1, United Kingdom]

Clinician Expectations

Mental Health Staff May Act as a Filter

As seen in the data from both the team meeting notes and focus groups, the research team was concerned that mental health staff sometimes acted as gatekeepers for some service users. This *gate keeping* behavior appeared to be expressed when mental health staff assumed a potential participant would be unable to take part in the study, resulting in a filtering effect that biases which participants are invited to take part. Trial staff constructed that the concept of gatekeeping extended beyond participating in clinical research and was perhaps linked to

mental health staff feeling protective over patients in their caseload. In the following example, a researcher reflects on how mental health staff appeared to very quickly decide whether a service user could cope with the intervention:

> *Even when you approached them with eligible participants, they [staff] were maybe more likely to discount them straight away. Just say "no, they're not suitable," or "I don't think they want to take part."* [Participant, United Kingdom]

Mental Health Staffs' Resistance to Research Participation

Research staff working on EMPOWER theorized that mental health staffs' resistance to research participation emerged because mental health staff believed that they were expected to participate in clinical research as part of their role as mental health clinicians. There were some concerns that if mental health staff felt that their participation in the project was mandatory, this may have limited their motivation and commitment, resulting in resistance to participation. In the following example, a member of the EMPOWER trial reflects on an encounter with a clinician who stated that they had to become involved because of expectations from management. This appeared to be linked with hierarchical relationships within mental health services. Therefore, clinical staff participating within research appeared to be a role expectation for clinical staff:

> *I remember one staff member talking about whether he agreed to be involved and he said "oh, do I really have a choice?" kind of saying "well, we've heard about it from, you know, management" and I got the sense he was communicating there was an expectation to get involved but that was just one thing I picked up about that kind of involvement. Yeah.* [Participant, Australia]

Trial Work Used to Facilitate Recruitment

Trial staff used several trial work strategies to facilitate recruitment in face of barriers, including flexibility in approach to barriers, persistence, and emotional labor (trial staff managing feelings and expressions to successfully recruit participants), in addition to building relationships (using preexisting relationships with clinicians and using supportive research team relationships).

Flexibility in Approach to Barriers

Regardless of how barriers to recruitment were negotiated, something that stood out in both the minutes and the focus groups was the need for trial staff to be flexible in their approaches. Discussions around the benefits of the flexible approach were common throughout both the Australian and UK focus groups. In the following example, a team member from Australia highlights that being flexible (and not rigid) in their approach to recruitment enabled staff to work through problems as they occurred:

> *I think that one of the real strengths in our research team has been how flexible and adaptive we've been when these challenges have come up, everyone involved in the process has been really thinking about ways to problem solve these things and coming up with suggestions.* [Participant, Australia]

One example trial staff provided, which illustrates taking a flexible approach, was in their discussions with clinical staff surrounding the trial protocol. Within a feasibility study, information about the recruitment process is a key outcome. Therefore, when encountering potential staff *paternalism* toward patients on their caseload, trial staff could emphasize that knowing how many people would refuse to take part was an important trial outcome. Explaining to trial staff that the protocol required that all relevant participants should have the opportunity to be approached, to discover the number of patients who did not want to take part, was described as it could circumnavigate the perceived filtering behaviors by clinical staff. In the following example, a principal investigator also describes how being flexible could enable trial staff to resist or negotiate staff paternalism, without it seeming like a direct challenge to clinical judgment:

> *...and our primary method of trying to get around that was to blame a third party to blame the protocol which says we needed to screen everyone and invite everyone rather than, you know directly, it feeling [sic] more like a direct challenge to the judgement of the key clinicians.* [Participant, Australia]

The researcher noted in their reflective memo that flexibility appeared a key process that emerged from the very beginning of recruitment when trial staff were working to build relationships and engage with the staff. Trial staff did not appear to rigidly stick to one recruitment approach:

> *When looking through minutes from the start of the trial. I am struck by how apparent flexibility was from the early stages of recruitment. For example, working around the availability of clinical staff as much as was possible. Furthermore, it feels important to note that because clinical staff are so busy that being flexible appeared essential in moving recruitment forward. However, in later stages flexibility involved clinical trial staff.* [Researcher's reflective memo]

Persistence

Within EMPOWER, *trial work* was characterized not only by flexibility but also by persistence. This could be seen in accounts of trial staff constantly trying to contact mental health staff. The practical work of chasing up mental health staff was readily apparent from the analysis of the minutes of meetings and reflective accounts of the recruitment process recorded in both focus groups. Chasing up could involve telephone calls, email, or visits in person to community mental health teams. This was often because of systematic issues such as a lack of staff time to support the intervention but could also be because of local factors such as mental health staff feeling pressurized into taking part by management and resisting participation. However, linked to staff describing their need to be persistent, there was acknowledgment that chasing up mental health staff could be a time-consuming part of trial work:

> *It depended quite a lot on the key workers that were involved within teams. How open they were to the*

> *study, and how much they followed through on things they said they were going to do. So, a lot of the time was spent chasing up key workers who said they would do something, and then didn't.* [Participant, United Kingdom]

Emotional Labor

Although the need to be persistent in chasing up mental health staff and trying out different recruitment strategies was apparent from both the minutes of meetings and the focus groups, the focus groups foregrounded an important role for the emotional aspects of recruitment within a clinical trial. In the following example, it is clear that simply being persistent is not enough and that it is important for it not to be obvious that the research team experienced frustration. Indeed, the need to portray constant positivity to get the work done appeared to be considered key in successfully recruiting participants. Therefore, there appeared to be an important role for *emotional labor* within trial work:

> *Persistence. Always smiling. Always the utmost professionalism.* [Participant 1]
>
> *Sometimes it's fake. [shared laughter].* [Participant 6, United Kingdom]
>
> *To the best of my knowledge, no trial staff used the term emotional labour to describe the maintaining professionalism during interactions with mental health staff, carers and patients. However, when reflecting on my observations of the research process, emotional labour appeared a highly relevant interactional framework for understanding the actual work underpinning trial staff describing the competency of staying polite and professional even when faced with potentially stressful challenges. Emotional labour seemed especially pertinent because trial staff are trying to invoke positive feelings within clinical research staff to build trust in both the project and the research team themselves.* [Researcher's reflective memo]

Relationship Building With Mental Health Staff

Trial work appeared to be sustained and facilitated by relationship building. When trial staff described the work that they performed throughout the recruitment process, at all stages, the work appeared to be underpinned by trial staffs' ability to successfully build and use relationships. In the absence of the ability to tap into existing relationships, trial staff had to be able to quickly build working relationships with clinical staff to facilitate the recruitment process. Reflecting on the overall emergent process, trial staff centered on the importance of building relationships with clinical staff in both the United Kingdom and Australia. One key change that came from this was trial staff becoming trusted to make direct approaches to patients instead of always having to go through mental health staff:

> *I think the reason that it became more possible was um that the services got used to the research team and got confident in the research team, or at least management did, so I think there's something about us building the relationship that enabled us to move into a different way of doing it.* [Participant, Australia]

By appraising the minutes of team meetings, it is clear that trial staff initially had to go almost entirely through mental health staff. However, if a good relationship was built, this was perceived as helpful for recruitment because the staff were generally more engaged with the team:

> *Within two months, trial work moved on to the establishment of relationships between mental health staff and the research team. In this stage, the EMPOWER staff became trusted to make direct approaches. Linked to the process of building relationships over time with mental health staff, in both Glasgow and Melbourne, a clinical team member [Research Nurse and Peer Support Worker, respectively] became involved in trial recruitment. Both teams reflected upon this positively because both of these clinical team members brought their pre-existing relationships with clinical staff. While the earlier stages of recruitment may have seemed slow, it appears productive in terms of carrying out trial work that built relationships and trust with clinical staff, ultimately moving trial recruitment forward.* [Researcher's reflective memo]

Using Preexisting Relationships

Although building relationships underpinned all aspects of trial work, preexisting relationships were described as helpful in establishing clinician trust. The *trial work* here is the insight and ability of the trial staff to use those preexisting relationships in the service of recruitment. In the example given below, a research assistant stated that clinical staff felt more comfortable communicating negative feelings about the recruitment process to the peer support worker (part of the EMPOWER trial team) because of preexisting ease and trust that come with already knowing someone. The research team was then able to use this information and adapt the approach taken to recruitment to be less aversive for clinical staff:

> *I think the real turning point where [peer support worker who participated in recruitment process] was speaking to somebody perhaps because she has that more casual kind of pre-existing relationship with some of these people where they were explicitly saying "I'm a bit sick of this EMPOWER stuff" and that's when you know, that sent out the message we need to pump the brakes hard in terms of how much we are asking clinicians to do here.* [Participant, Australia]

Relationship Building—Internal Within the Research Team

Relationships appeared to serve important internal functions within the EMPOWER team. Across both the United Kingdom and Australia, trial staff made reference to the importance of having team members who understood the challenges associated with clinical trial recruitment. Furthermore, the importance of having space to be open about difficulties encountered, so that discussions were focused on how best to move forward, was described:

> *Because I think at times it is quite demotivating. And particularly if you've got that third [unanswered] phone call and think "please just answer the phone." I think we [trial recruitment staff] do try and support each other through those times.* [Participant, United Kingdom]

> *From the meeting minutes, being part of the UK meetings while recruitment was on-going and appraising themes constructed during the focus groups, it seemed as though having a space within the trial team to discuss and share frustrations that were inevitable from negotiating the various recruitment barriers. From my observations of actual meetings and continued within the focus groups, there appeared to be lots of in-jokes within the teams about the recruitment process including challenging aspects. For trial staff, this appeared to provide camaraderie and support.* [Researcher's reflective memo]

To summarize, relationship building internally within the team appeared to be just as important in facilitating the recruitment process as building external relationships with mental health staff. Trial staff were there for each other throughout recruitment challenges and provided a supportive space for each other to discuss problems.

Discussion

Principal Findings

This study explored recruitment from the point of view of trial staff working on a digital intervention for psychosis. Detailed descriptions of the recruitment process are rarely reported within RCTs of digital interventions for psychosis, which minimizes the opportunity for sharing learning on how best to overcome recruitment barriers. By examining the recruitment process in EMPOWER using ethnography supplemented with focus groups, we now present such a detailed description. In doing so, we demonstrate not only the kind of recruitment barriers encountered by trial staff but also what strategies trial staff use to overcome them. Recruitment barriers appeared to span macro (structure and systems, eg, lack of staff time), meso (roles, eg, staff leadership), and micro (idiosyncratic, eg, the physical layout of community mental health premises) levels. The findings from this qualitative study suggest that simply reporting the number of participants recruited (n=86) clouds a highly complex social process underpinning trial recruitment. Taken together, the findings from this study can start to theorize the recruitment barriers and facilitators within the recruitment process for the EMPOWER trial.

Although it has been recommended that research exploring recruitment barriers should go beyond reporting a lack of staff time [31], it appeared a systemic problem within this trial that trial staff found difficult to negotiate. Lack of staff time has been reported as a recruitment challenge in many mental health studies [56]. Therefore, our results support those of Skea et al [54], who suggested that researchers should consider how essential trial recruitment processes fit in with the reality of clinical practice. The nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [57] provides a framework for understanding challenges encountered in the implementation of digital technologies. NASSS frames challenges as being simple (straightforward and predictable), complicated (multiple interacting components), or complex (unpredictable and hard to reduce down into linear components). NASSS addresses challenges and complexities that occur in different domains when implementing health care technologies, including the health condition being intervened on, value proposition, technology, adopter system, organization, wider social context, and changes over time. When framing the recruitment process via health care organizations in the United Kingdom and Australia, it appears that macrolevel recruitment barriers pose particularly complex challenges because of severe resource pressures, with staff struggling to find time to support research, as noted by other clinical trial researchers [58]. However, even more idiosyncratic challenges such as differences in leadership between cluster sites were noted by trial staff to have complex, unpredictable, and sometimes large impacts on recruitment, supporting the need to understand contextual differences across clusters in cRCTs [39].

To negotiate complex recruitment barriers, trial staff put significant amounts of work in to engaging mental health staff during the recruitment process. Trial work is multifactorial and comprises emotional labor and social and professional competencies. Initially, in performing trial work, staff in EMPOWER reported the importance of persistence, being flexible in trying different approaches, and always being professional in their interactions with staff. Previous research on clinical trial staff has suggested that emotional labor is a key part of trial work when staff are working to meet recruitment targets [59]. In the face of stresses and strains created by recruitment barriers, trial staff have a duty to maintain an ethos of professionalism. Coming from the field of sociology, emotional labor is described as the silent work of evoking feelings in others and managing one's own emotional expressions to do so [60]. Emotional labor appeared a key strategy when dealing with barriers such as having to pursue contact with very busy staff while maintaining good working relationships by not letting frustrations show. Relationships between trial staff and clinicians (and the ability to quickly build and rapport) appeared essential to successful recruitment. However, barriers existed in the recruitment process, which could make relationship building difficult. Although a lack of clinical staff time is well reported in the literature, factors such as the layout of buildings, making it impossible to have a private conversation, also acted as a relationship building block.

Clinicians' exclusion of people independent of trial protocol criteria is noted to be a key challenge in mental health intervention recruitment [56,61]. In the case of EMPOWER, it appeared that clinicians regularly sought to exclude participants for reasons not stated in the protocol. Trial staff were given the impression that this was because of clinical staff having concerns about a service user's ability to cope with study participation. However, trial staff sometimes seemed able to negotiate this challenge by invoking the trial protocol and reminding staff that determining directly from service users their willingness (or not) to participate was an important outcome within a feasibility study. Mental health staff filtering what patients ended up being

approached for recruitment was a key theme identified in previous research exploring barriers to recruitment to nondigital psychosis studies [62]. Excluding participants for reasons not contained in the protocol likely has implications for the replicability and robustness of research findings because the selection criteria are obscured [61] and samples likely become biased. Therefore, there is a need to learn more about why this apparent *filtering* happens (from the perspective of mental health staff), particularly in digital interventions for psychosis where little is currently known [13] and there may be assumptions about ability of people with psychosis to use technology [12].

Mental health staff have perceptions of what is required from them professionally, and these perceptions seemed to cause tension and role conflict during the recruitment process. For example, clinical staff may not feel that they have the autonomy to decline participation because participating in research is a role expectation for clinical staff. Previous oncology research has indicated that nurses involved in conducting research describe a role conflict, where duty of care to the patient can sit uncomfortably with *feeling like a salesperson* when encouraging patient participation within trials [63]. Enhancing collaborations with key stakeholders such as mental health staff is stated to be important in developing better digital interventions for psychosis [20]. Therefore, it seems pertinent to understand issues such as role conflict from the perspective of trial staff and co-design recruitment procedures around the needs of mental health staff.

Persistence and flexibility of approach were important in negotiating everything from macrolevel barriers, such as a lack of staff time, to more microlevel issues, such as community mental health center managers with different styles. One key element of the flexible approach to recruitment that emerged during the EMPOWER trial was a peer support worker (a person with their own experiences of psychosis employed to support people in their use of the intervention) advising how to approach recruitment challenges. A review concluded that patient involvement in clinical research may be associated with increased recruitment (but not retention) in clinical trials [64]. However, the mechanisms underlying this effect are still unclear. Within EMPOWER, actively transforming the peer support role to encompass involvement in recruitment was reported by trial staff to have been very useful for recruitment because the peer support role brought preexisting relationships with staff and fresh insight on how best to approach recruitment challenges. Although this may be very specific to EMPOWER, it nonetheless demonstrates that experiential knowledge and enhanced capacity for relationship building with clinical staff may be important mechanisms to consider when theorizing mechanisms of patient and public involvement in trial recruitment.

Future Research

The research team reported that conveying to staff that highlighting the importance of gathering data on rates of participant refusal was helpful in negotiating filtering behavior by clinical staff. Future research could explore this observed phenomenon further, perhaps using relevant behavioral change theories as a theoretical framework [65]. Emotional labor in the context of clinical trials has previously been theorized in recruitment research involving direct interaction with patients [59]. However, these findings suggest that emotional labor may be relevant in the everyday work of keeping clinical staff engaged in the recruitment process. The EMPOWER trial was conducted simultaneously in Australia and the United Kingdom. Therefore, it is perhaps unsurprising that a specific recruitment issue unique to one health care system that was observed (high patient turnover within Australia) was apparent. However, there were some marked similarities across countries, such as a lack of staff time. Clinical trials conducted across multiple countries may benefit from providing some context on differences between mental health care systems to contextualize recruitment results. In addition, a Delphi study [66] could expand upon the barriers identified here to see if they are more widespread in trials of similar interventions.

Limitations

EMPOWER was a feasibility study, which means there were a limited number of trial staff to observe and speak to. Beyond the small sample, the findings from this study should be considered in light of several key limitations. Ethnography is an opportunistic methodology [67]; therefore, researchers are limited by what they can or are allowed to observe. With regard to research methods, we did not believe that the focus group conducted remotely was any less rich than the focus group that was conducted in person in terms of the transcripts produced. However, it is important to highlight that conducting one focus group in person and another remotely may have impacted both the conduct of the research and the analysis. Moreover, although Australian recruitment was discussed at UK-based meetings and was recorded in the minutes there, SA did not attend any Australian recruitment meetings because of being based in the United Kingdom and did not directly observe Australian staff during the recruitment process. Although this study identified barriers and suggested potential ways to optimize recruitment, the potential positive impact of qualitative research in trial recruitment research needs to be further explored [35] before any comment can be made about potential utility. Furthermore, we have not focused on retention, which is also an important issue in its own right [68,69]. In addition, this study focused on barriers and facilitators experienced by trial staff during the recruitment phase of the trial and are likely biased toward their own perspectives.

Facilitators addressing ongoing *service characteristics* such as staff turnover and physical environment may have emerged if the study had been widened to include service managers or other informants. Furthermore, there was not much focus on the experiences of service user participants throughout the focus groups. Future research understanding barriers and facilitators to recruitment from the point of view of service users within clinical trials of digital interventions for psychosis, building upon previous work exploring what service users think about digital interventions for psychosis in general [70-72] and their feelings about recruitment into a clinical trial for distressing voices that involved interacting with a digital avatar [72]. Another key limitation is that recruitment within EMPOWER occurred in public mental health care systems in both Australia and the United Kingdom; however, recruitment in private health

care systems or recruitment processes conducted remotely through the internet may have unique challenges. Finally, the focus of this study was to empirically explore recruitment from the point of view of trial staff; however, it is important to highlight that future research would benefit from exploring recruitment from the perspectives of clinical staff and service users, which will develop a more ecologically valid overview of the recruitment process.

Conclusions

Exploring recruitment from the perspective of trial staff provides rich insights into barriers and facilitators to recruitment within clinical trials of digital intervention. For example, rather than people diagnosed with schizophrenia being a *hard-to-reach group*, it seems that difficulties in recruiting people diagnosed with schizophrenia to clinical trials emerge from complex dynamic interactions within health care systems. This study suggests that recruitment in a clinical trial of a digital intervention for psychosis is complex. Barriers to recruitment exist at micro, meso, and macro levels, and trial staff must negotiate these barriers within their role to meet recruitment targets to the best of their abilities. Key competencies observed during the recruitment process included flexibility, persistence, and emotional labor. As discussed in focus groups and aligned with ethnographic observations, it was important for trial staff to work within a team that understood that recruitment to clinical trials could be challenging and appreciated having access to peer support from other trial staff. People responsible for managing staff who recruit into clinical trials may wish to consider these relationship-focused factors when deciding how best to supervise staff and design effective and resilient teams. One key conclusion from this study is that learning about what works along the way is important, as it provides a space for trial staff to discuss the recruitment process and both learn from and support each other during recruitment. Relationship building with clinical staff appeared to help facilitate the recruitment process, which may have important implications for credentialing, training, and supervising staff who work within clinical trials.

Acknowledgments

The authors are grateful to all the service users, carers and mental health staff, and CMHS who gave their time and resources to contribute to the development of the EMPOWER study during the consultation phase and to the c-RCT and the process evaluation. This was critical in developing the process evaluation.

Conflicts of Interest

None declared.

References

1. Kannisto KA, Korhonen J, Adams CE, Koivunen MH, Vahlberg T, Välimäki MA. Factors Associated With Dropout During Recruitment and Follow-Up Periods of a mHealth-Based Randomized Controlled Trial for Mobile.Net to Encourage Treatment Adherence for People With Serious Mental Health Problems. J Med Internet Res 2017 Feb 21;19(2):e46 [FREE Full text] [doi: 10.2196/jmir.6417] [Medline: 28223262]
2. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med 2018 Aug;210:2-21 [FREE Full text] [doi: 10.1016/j.socscimed.2017.12.005] [Medline: 29331519]
3. Tudur Smith C, Hickey H, Clarke M, Blazeby J, Williamson P. The trials methodological research agenda: results from a priority setting exercise. Trials 2014 Jan 23;15:32 [FREE Full text] [doi: 10.1186/1745-6215-15-32] [Medline: 24456928]
4. Walters SJ, Bonacho Dos Anjos Henriques-Cadby I, Bortolami O, Flight L, Hind D, Jacques RM, et al. Recruitment and retention of participants in randomised controlled trials: a review of trials funded and published by the United Kingdom Health Technology Assessment Programme. BMJ Open 2017 Mar 20;7(3):e015276 [FREE Full text] [doi: 10.1136/bmjopen-2016-015276] [Medline: 28320800]
5. Liu Y, Pencheon E, Hunter RM, Moncrieff J, Freemantle N. Recruitment and retention strategies in mental health trials - a systematic review. PLoS One 2018 Aug 29;13(8):e0203127 [FREE Full text] [doi: 10.1371/journal.pone.0203127] [Medline: 30157250]
6. Halpern SD, Karlawish JHT, Berlin JA. The continuing unethical conduct of underpowered clinical trials. JAMA 2002 Jul 17;288(3):358-362. [doi: 10.1001/jama.288.3.358] [Medline: 12117401]
7. Heller C, Balls-Berry JE, Nery JD, Erwin PJ, Littleton D, Kim M, et al. Strategies addressing barriers to clinical trial enrollment of underrepresented populations: a systematic review. Contemp Clin Trials 2014 Nov;39(2):169-182 [FREE Full text] [doi: 10.1016/j.cct.2014.08.004] [Medline: 25131812]
8. Mills E, Wilson K, Rachlis B, Griffith L, Wu P, Guyatt G, et al. Barriers to participation in HIV drug trials: a systematic review. The Lancet Infectious Diseases 2006 Jan;6(1):32-38. [doi: 10.1016/s1473-3099(05)70324-8]
9. Fayter D, McDaid C, Eastwood A. A systematic review highlights threats to validity in studies of barriers to cancer trial participation. J Clin Epidemiol 2007 Oct;60(10):990-1001. [doi: 10.1016/j.jclinepi.2006.12.013] [Medline: 17884592]
10. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]

11. Orben A, Przybylski AK. The association between adolescent well-being and digital technology use. Nat Hum Behav 2019 Feb 14;3(2):173-182. [doi: 10.1038/s41562-018-0506-1] [Medline: 30944443]
12. Allan S, Bradstreet S, Mcleod H, Farhall J, Lambrou M, Gleeson J, et al. Developing a Hypothetical Implementation Framework of Expectations for Monitoring Early Signs of Psychosis Relapse Using a Mobile App: Qualitative Study. J Med Internet Res 2019 Oct 24;21(10):e14366. [doi: 10.2196/14366]
13. Aref-Adib G, McCloud T, Ross J, O'Hanlon P, Appleton V, Rowe S, et al. Factors affecting implementation of digital health interventions for people with psychosis or bipolar disorder, and their family and friends: a systematic review. The Lancet Psychiatry 2019 Mar;6(3):257-266 [FREE Full text] [doi: 10.1016/s2215-0366(18)30302-x]
14. Finley EP, Huynh AK, Farmer MM, Bean-Mayberry B, Moin T, Oishi SM, et al. Periodic reflections: a method of guided discussions for documenting implementation phenomena. BMC Med Res Methodol 2018 Nov 27;18(1):153 [FREE Full text] [doi: 10.1186/s12874-018-0610-y] [Medline: 30482159]
15. Cheng K, Metcalfe A. Qualitative Methods and Process Evaluation in Clinical Trials Context. International Journal of Qualitative Methods 2018 Apr 30;17(1) [FREE Full text] [doi: 10.1177/1609406918774212]
16. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: Medical Research Council guidance. BMJ 2015 Mar 19;350(mar19 6):h1258-h1258 [FREE Full text] [doi: 10.1136/bmj.h1258]
17. Moore G, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: a summary of Medical Research Council guidance. In: Richards D, Hallberg IR, editors. Complex interventions in health: an overview of research methods. Abingdon: Routledge; 2015:222.
18. 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 22;42(2):448-455 [FREE Full text] [doi: 10.1093/schbul/sbv132] [Medline: 26400871]
19. Bell IH, Lim MH, Rossell SL, Thomas N. Ecological Momentary Assessment and Intervention in the Treatment of Psychotic Disorders: A Systematic Review. Psychiatr Serv 2017 Nov 01;68(11):1172-1181. [doi: 10.1176/appi.ps.201600523] [Medline: 28669284]
20. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. Psychol Psychother 2019 Jun 28;92(2):277-297. [doi: 10.1111/papt.12222] [Medline: 30924316]
21. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
22. Lim M, Penn D. Using Digital Technology in the Treatment of Schizophrenia. Schizophr Bull 2018 Aug 20;44(5):937-938 [FREE Full text] [doi: 10.1093/schbul/sby081] [Medline: 29878251]
23. Torous J, Firth J. Bridging the dichotomy of actual versus aspirational digital health. World Psychiatry 2018 Feb;17(1):108-109 [FREE Full text] [doi: 10.1002/wps.20464] [Medline: 29352531]
24. Jørgensen R, Munk-Jørgensen P, Lysaker PH, Buck KD, Hansson L, Zoffmann V. Overcoming recruitment barriers revealed high readiness to participate and low dropout rate among people with schizophrenia in a randomized controlled trial testing the effect of a Guided Self-Determination intervention. BMC Psychiatry 2014 Feb 3;14(1). [doi: 10.1186/1471-244x-14-28]
25. Carmichael J, Misselbrook T, Marriner L, Forrest A, Molodynski A. Clinicians' attitudes towards patient involvement in mental health research. Prog Neurol Psychiatry 2016 Feb 09;20(1):22-25. [doi: 10.1002/pnp.413]
26. Agarwal S, LeFevre AE, Lee J, L'Engle K, Mehl G, Sinha C, WHO mHealth Technical Evidence Review Group. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016 Mar 17;352:i1174. [doi: 10.1136/bmj.i1174] [Medline: 26988021]
27. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. BMC Med Res Methodol 2010;10:67 [FREE Full text] [doi: 10.1186/1471-2288-10-67] [Medline: 20637084]
28. Campbell M, Piaggio G, Elbourne D, Altman D, CONSORT Group. Consort 2010 statement: extension to cluster randomised trials. BMJ 2012 Sep 04;345(7881):e5661. [doi: 10.1136/bmj.e5661] [Medline: 22951546]
29. Kearney A, Harman N, Rosala-Hallas A, Beecher C, Blazeby J, Bower P, et al. Development of an online resource for recruitment research in clinical trials to organise and map current literature. Clin Trials 2018 Dec;15(6):533-542 [FREE Full text] [doi: 10.1177/1740774518796156] [Medline: 30165760]
30. Montgomery P, Grant S, Mayo-Wilson E, Macdonald G, Michie S, Hopewell S, CONSORT-SPI Group. Reporting randomised trials of social and psychological interventions: the CONSORT-SPI 2018 Extension. Trials 2018 Jul 31;19(1):407 [FREE Full text] [doi: 10.1186/s13063-018-2733-1] [Medline: 30060754]
31. Glasgow RE, Huebschmann AG, Brownson RC. Expanding the CONSORT Figure: Increasing Transparency in Reporting on External Validity. Am J Prev Med 2018 Sep;55(3):422-430. [doi: 10.1016/j.amepre.2018.04.044] [Medline: 30033029]
32. Harris J, Booth A, Cargo M, Hannes K, Harden A, Flemming K, et al. Cochrane Qualitative and Implementation Methods Group guidance series-paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. J Clin Epidemiol 2018 May;97:39-48 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.10.023] [Medline: 29248725]

33. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials 2013 Jan 12;14:15 [FREE Full text] [doi: 10.1186/1745-6215-14-15] [Medline: 23311722]
34. Druce K, McBeth J, van der Veer SN, Selby D, 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 Nov 01;5(11):e168 [FREE Full text] [doi: 10.2196/mhealth.8162] [Medline: 29092810]
35. Hennessy M, Hunter A, Healy P, Galvin S, Houghton C. Improving trial recruitment processes: how qualitative methodologies can be used to address the top 10 research priorities identified within the PRioRiTy study. Trials 2018 Oct 25;19(1):584 [FREE Full text] [doi: 10.1186/s13063-018-2964-1] [Medline: 30359293]
36. O'Cathain A, Thomas K, Drabble S, Rudolph A, Hewison J. What can qualitative research do for randomised controlled trials? A systematic mapping review. BMJ Open 2013 Jun 20;3(6) [FREE Full text] [doi: 10.1136/bmjopen-2013-002889] [Medline: 23794542]
37. Treweek S, Bevan S, Bower P, Campbell M, Christie J, Clarke M, et al. Trial Forge Guidance 1: what is a Study Within A Trial (SWAT)? Trials 2018 Feb 23;19(1):139 [FREE Full text] [doi: 10.1186/s13063-018-2535-5] [Medline: 29475444]
38. Gumley A, Bradstreet S, Ainsworth J, Allan S, Alvarez-Jimenez M, Beattie L, et al. Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-Being, Engagement, and Recovery: Protocol for a Feasibility Cluster Randomized Controlled Trial Harnessing Mobile Phone Technology Blended With Peer Support. JMIR Res Protoc 2020 Jan 09;9(1):e15058 [FREE Full text] [doi: 10.2196/15058] [Medline: 31917372]
39. Mann C, Shaw A, Guthrie B, Wye L, Man M, Hollinghurst S, et al. Protocol for a process evaluation of a cluster randomised controlled trial to improve management of multimorbidity in general practice: the 3D study. BMJ Open 2016 May 04;6(5):e011260 [FREE Full text] [doi: 10.1136/bmjopen-2016-011260] [Medline: 27147391]
40. Allan S, Mcleod H, Bradstreet S, Beedie S, Moir B, Gleeson J, et al. Understanding Implementation of a Digital Self-Monitoring Intervention for Relapse Prevention in Psychosis: Protocol for a Mixed Method Process Evaluation. JMIR Res Protocols 2019 Dec 10;8(12):e15634 [FREE Full text] [doi: 10.2196/15634] [Medline: 31821154]
41. Thirsk L, Clark A. Using Qualitative Research for Complex Interventions. International Journal of Qualitative Methods 2017 Aug 24;16(1):160940691772106. [doi: 10.1177/1609406917721068]
42. Snowdon C. Qualitative and mixed methods research in trials. Trials 2015 Dec 08;16(1):558 [FREE Full text] [doi: 10.1186/s13063-015-1084-4] [Medline: 26646545]
43. Palinkas LA, Zatzick D. Rapid Assessment Procedure Informed Clinical Ethnography (RAPICE) in Pragmatic Clinical Trials of Mental Health Services Implementation: Methods and Applied Case Study. Adm Policy Ment Health 2019 Mar 28;46(2):255-270 [FREE Full text] [doi: 10.1007/s10488-018-0909-3] [Medline: 30488143]
44. Reynolds J. 'Missing out': Reflections on the positioning of ethnographic research within an evaluative framing. Ethnography 2016 Aug 17;18(3):345-365. [doi: 10.1177/1466138116664106]
45. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Jun 20;16(1):95 [FREE Full text] [doi: 10.1186/s12916-018-1089-4] [Medline: 29921272]
46. Morgan-Trimmer S, Wood F. Ethnographic methods for process evaluations of complex health behaviour interventions. Trials 2016 May 04;17(1):232 [FREE Full text] [doi: 10.1186/s13063-016-1340-2] [Medline: 27142662]
47. Murdoch J. Process evaluation for complex interventions in health services research: analysing context, text trajectories and disruptions. BMC Health Serv Res 2016 Aug 19;16(1):407 [FREE Full text] [doi: 10.1186/s12913-016-1651-8] [Medline: 27538946]
48. Kitchen C, Lewis S, Tiffin P, Welsh P, Howey L, Ekers D. A focused ethnography of a Child and Adolescent Mental Health Service: factors relevant to the implementation of a depression trial. Trials 2017 May 25;18(1):237 [FREE Full text] [doi: 10.1186/s13063-017-1982-8] [Medline: 28545545]
49. Rick J, Clarke M, Montgomery AA, Brocklehurst P, Evans R, Bower P. Doing trials within trials: a qualitative study of stakeholder views on barriers and facilitators to the routine adoption of methodology research in clinical trials. Trials 2018 Sep 10;19(1):481-416 [FREE Full text] [doi: 10.1186/s13063-018-2862-6] [Medline: 30201022]
50. NVivo Qualitative Data Analysis Software Version 11. QSR International. 2015. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2020-01-08]
51. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
52. Charmaz K. Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. London, UK: SAGE Publications; 2006.
53. Doyle S. Member checking with older women: a framework for negotiating meaning. Health Care Women Int 2007;28(10):888-908. [doi: 10.1080/07399330701615325] [Medline: 17987459]
54. Skea Z, Treweek S, Gillies K. 'It's trying to manage the work': a qualitative evaluation of recruitment processes within a UK multicentre trial. BMJ Open 2017 Aug 11;7(8):e016475 [FREE Full text] [doi: 10.1136/bmjopen-2017-016475] [Medline: 28801422]

55. Gale N, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13(1):117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
56. Jones H, Cipriani A. Barriers and incentives to recruitment in mental health clinical trials. Evid Based Ment Health 2019 May;22(2):49-50. [doi: 10.1136/ebmental-2019-300090] [Medline: 31023822]
57. 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]
58. Sheard L, Peacock R. Fiddling while Rome burns? Conducting research with healthcare staff when the NHS is in crisis. JHOM 2019 Dec 09;34(1):77-84. [doi: 10.1108/jhom-04-2019-0105]
59. Lawton J, Kirkham J, White D, Rankin D, Cooper C, Heller S. Uncovering the emotional aspects of working on a clinical trial: a qualitative study of the experiences and views of staff involved in a type 1 diabetes trial. Trials 2015 Jan 07;16(1):3 [FREE Full text] [doi: 10.1186/1745-6215-16-3] [Medline: 25566971]
60. Abstract NJ. Emotional Labour: Skill and Work in the Social Regulation of Feelings. The Sociological Review 2011 Aug 04;37(1):15-42. [doi: 10.1111/j.1467-954x.1989.tb00019.x]
61. Pinfold V, Cotney J, Hamilton S, Weeks C, Corker E, Evans-Lacko S, et al. Improving recruitment to healthcare research studies: clinician judgements explored for opting mental health service users out of the time to change viewpoint survey. J Ment Health 2019 Feb;28(1):42-48 [FREE Full text] [doi: 10.1080/09638237.2017.1340598] [Medline: 28699815]
62. Bucci S, Butcher I, Hartley S, Neil S, Mulligan J, Haddock G. Barriers and facilitators to recruitment in mental health services: care coordinators' expectations and experience of referring to a psychosis research trial. Psychol Psychother 2015 Sep;88(3):335-350. [doi: 10.1111/papt.12042] [Medline: 25257960]
63. Tinkler L, Smith V, Yiannakou Y, Robinson L. Professional identity and the Clinical Research Nurse: A qualitative study exploring issues having an impact on participant recruitment in research. J Adv Nurs 2018 Feb;74(2):318-328. [doi: 10.1111/jan.13409] [Medline: 28792610]
64. Crocker J, Hughes-Morley A, Petit-Zeman S, Rees S. Assessing the impact of patient and public involvement on recruitment and retention in clinical trials: a systematic review. Trials 2015 Nov 16;16(S2):a [FREE Full text] [doi: 10.1186/1745-6215-16-s2-o91]
65. Musker M, Short C, Licinio J, Wong M, Bidargaddi N. Using behaviour change theory to inform an innovative digital recruitment strategy in a mental health research setting. J Psychiatr Res 2020 Jan;120:1-13 [FREE Full text] [doi: 10.1016/j.jpsychires.2019.09.012] [Medline: 31610405]
66. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. J Adv Nurs 2000 Oct;32(4):1008-1015. [Medline: 11095242]
67. Conte K, Shahid A, Grøn S, Loblay V, Green A, Innes-Hughes C, et al. Capturing implementation knowledge: applying focused ethnography to study how implementers generate and manage knowledge in the scale-up of obesity prevention programs. Implement Sci 2019 Sep 18;14(1):91 [FREE Full text] [doi: 10.1186/s13012-019-0938-7] [Medline: 31533765]
68. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
69. Skea Z, Newlands R, Gillies K. Exploring non-retention in clinical trials: a meta-ethnographic synthesis of studies reporting participant reasons for drop out. BMJ Open 2019 Jun 03;9(6):e021959 [FREE Full text] [doi: 10.1136/bmjopen-2018-021959] [Medline: 31164359]
70. 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 [FREE Full text] [doi: 10.1186/s12888-018-1979-1] [Medline: 30665384]
71. Eisner E, Drake R, Berry N, Barrowclough C, Emsley R, Machin M, et al. Development and Long-Term Acceptability of ExPRESS, a Mobile Phone App to Monitor Basic Symptoms and Early Signs of Psychosis Relapse. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e11568 [FREE Full text] [doi: 10.2196/11568] [Medline: 30924789]
72. Hall J, Rus-Calafell M, Omari-Asor L, Ward T, Emsley R, Garety P, et al. Assessing the subjective experience of participating in a clinical trial (AVATAR). Psychiatry Res 2018 May;263:82-87 [FREE Full text] [doi: 10.1016/j.psychres.2018.02.026] [Medline: 29502043]

Abbreviations

CMHS: community mental health services
CONSORT: Consolidated Standards of Reporting Trials
cRCT: cluster randomized control trial
EMPOWER: Early Signs Monitoring to Prevent Relapse in Psychosis and Promote Well-being, Engagement, and Recovery
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability

RCT: randomized control trial
SWAT: study within a trial

Edited by A Kushniruk; submitted 02.09.20; peer-reviewed by A Hudon, S Byrne; comments to author 28.09.20; revised version received 15.11.20; accepted 23.12.20; published 05.03.21.

Please cite as:
Allan S, Mcleod H, Bradstreet S, Bell I, Whitehill H, Wilson-Kay A, Clark A, Matrunola C, Morton E, Farhall J, Gleeson J, Gumley A
Perspectives of Trial Staff on the Barriers to Recruitment in a Digital Intervention for Psychosis and How to Work Around Them: Qualitative Study Within a Trial
JMIR Hum Factors 2021;8(1):e24055
URL: https://humanfactors.jmir.org/2021/1/e24055
doi:10.2196/24055
PMID:33666555

Original Paper

Information Sources, Risk Perception, and Efficacy Appraisal's Prediction of Engagement in Protective Behaviors Against COVID-19 in China: Repeated Cross-sectional Survey

Jian Raymond Rui[1*], PhD; Keqing Yang[1*], PhD; Juan Chen[1*], PhD

College of Journalism and Communication, South China University of Technology, Guangzhou, China

[*]all authors contributed equally

Corresponding Author:
Keqing Yang, PhD
College of Journalism and Communication
South China University of Technology
382 Waihuan East Rd
Guangzhou, 510006
China
Phone: 86 17819720631
Email: jckqyang@scut.edu.cn

Abstract

Background: As the COVID-19 pandemic has become a major public health threat worldwide, it is critical to understand what factors affect individual engagement in protective actions. Because of its authoritarian political system and state-owned media system, how Chinese individuals engaged in protective actions against COVID-19 might be different compared to other countries.

Objective: The purpose of this study is to examine how the source of information about COVID-19, Chinese individuals' risk perception of COVID-19 (ie, perceived severity and perceived susceptibility), and their efficacy appraisal in controlling COVID-19 (ie, response efficacy and self-efficacy) affected their engagement in protective actions. Additionally, this study aims to investigate whether there is any difference in these relationships throughout the duration of this pandemic.

Methods: A six-wave repeated cross-sectional survey (N=1942) was conducted in six major cities in China between February 7 and April 23, 2020. Participants' reliance on expert versus inexpert sources for information about COVID-19, their perceived severity of and susceptibility to COVID-19, their response efficacy and self-efficacy, and their engagement in protective actions (staying at home, wearing a face mask, and washing hands) were measured. Demographic variables (sex, age, income, education, and city of residence), knowledge of COVID-19, and self-rated health condition were controlled.

Results: Reliance on expert sources did not become the major factor that motivated these actions until wave 3, and the negative effect of inexpert sources on these actions was limited to wave 2. Perceived severity encouraged some protective behaviors but its effect varied depending on the specific behavior. In addition, perceived severity exhibited a stronger effect on these behaviors compared to perceived susceptibility. The positive effect of response efficacy was only significant at waves 1 and 2, and limited to certain behaviors.

Conclusions: Chinese individuals' engagement in protective behaviors might not entirely be their autonomous decision but a result of compliance with executive orders. After the early outbreak, expert sources started to facilitate protective behaviors, suggesting that it might take time to develop trust in these sources. The facilitating effect of perceived severity lasted throughout the duration of the pandemic, but that of response efficacy was limited to the early stage.

(JMIR Hum Factors 2021;8(1):e23232) doi:10.2196/23232

KEYWORDS

information source; perceived severity; perceived susceptibility; response efficacy; self-efficacy; health information; protective behavior; COVID-19; protection; behavior; risk; perception; prediction

Introduction

Background

Having spread to 188 countries and regions [1], COVID-19 has become a serious public health threat worldwide. As of October 16, 2020, COVID-19 has caused over 38 million cases and nearly 1,100,000 deaths [1]. China is the first country where COVID-19 was discovered. As early as December 2019, COVID-19 was found in Wuhan, China [2]. The number of confirmed cases and deaths in China grew rapidly in January but started to decline in late February [3]. As of October 16, 2020, China reported 90,899 cases including 4739 deaths [1].

Despite scientific efforts, much about COVID-19 still remains uncertain, such as its origin and mutation [4]. Thus, given its high levels of risks, individuals are encouraged to take protective actions [5,6]. The extant research has explicated how individuals' engagement in protective behavior against COVID-19 varied depending on their knowledge, fear, risk perception, morality, and internet use [7-12].

However, empirical evidence in China is still scarce. The authoritarian political system in China reduced resistance to the government's executive orders such as locking down cities and placing citizens under quarantine [13], which controlled the spread of the pandemic [3]. In addition, the state ownership of media in China enables the government to provide large-scale health education and campaigns consistently, which might have facilitated engagement in protective actions. Thus, investigations on what factors affected Chinese individuals' engagement in protective actions against COVID-19 may provide additional knowledge on the potential influence of a unique sociocultural environment on health behavior.

However, to the best of our knowledge, only one study was conducted on how Chinese individuals performed protective actions against COVID-19 [12]. Furthermore, that study is a one-time cross-sectional investigation at the early stage of the outbreak [12]. As it remains unknown when the pandemic might end, it is critical to examine how factors related to taking preventive measures against COVID-19 might change across different stages of its outbreak. Therefore, this study employs a repeated cross-sectional approach to address this limitation. Specifically, built on the extended parallel process model (EPPM) [13], this study aims to test the theory by examining how individuals' risk perception of COVID-19 and their efficacy appraisal in controlling COVID-19 might affect their engagement in protective actions. Moreover, we seek to add to the extant research by investigating the role that one's reliance on different sources for information about COVID-19 plays in performing these protective actions. Although we built our study on EPPM, other theoretical work such as the protection motivation theory [14], the health belief model [15], and the risk perception attitude framework [16] also considered variables in EPPM and made similar predictions.

EPPM

EPPM contends that whether individuals engage in protective behaviors depends on their risk perception and efficacy appraisal [17]. Risk perception is usually conceptualized as the sum or average of perceived severity and perceived susceptibility [18]. Perceived severity refers to one's perception of the adversity of consequences if individuals do not engage in recommended actions, whereas perceived susceptibility is conceptualized as the likelihood that one is subject to the given health threat [18-20]. Additionally, efficacy appraisal is conceptualized as the sum of response efficacy and self-efficacy [18]. Response efficacy refers to the extent to which individuals think recommended protective actions can manage the given threat effectively, whereas self-efficacy is conceptualized as individuals' confidence in performing those recommended acts [17,18].

The original research on EPPM posits that whether risk perception may facilitate engagement in protective actions depends on the level of efficacy appraisal [17,21]. Specifically, risk perception can only motivate individuals to perform protective actions at high levels of efficacy appraisal, whereas this positive relationship is absent at low levels of efficacy appraisal [17,21]. However, subsequent work demonstrated that risk perception can drive protective actions without high efficacy appraisal [22] because the innate aversion to loss prompts individuals to avoid potential risks by taking preventive measures [23]. Therefore, higher levels of perceived severity and perceived susceptibility may be associated with heightened motivation to perform protective actions [22,24-28].

In addition, individuals reporting high levels of response efficacy are more driven to engage in behaviors that can minimize the threat [26,29,30] because this confidence is often correlated with enhanced levels of hope [31]. Moreover, individuals reporting high levels of self-efficacy are more likely to follow the recommended acts because they tend to think it is less challenging to perform those behaviors [32]. Taken together, we predict that response efficacy and self-efficacy in controlling COVID-19 should exhibit positive relationships with engagement in protective behaviors.

Information Sources About COVID-19

Individuals equipped with accurate health information are usually more motivated to engage in health behaviors [9]. However, the volume of rumors about COVID-19 makes individuals vulnerable to health misinformation [33]. One factor that could potentially affect the credibility of information is its source. We categorized information sources into expert versus inexpert sources. Expert sources are conceptualized as individuals with medical expertise and organizations with professional gatekeepers that can screen information before it is published. These expert sources include expert media, government administrations, expert health organizations, and medical experts. The gatekeeping theory contends that gatekeepers, or people screening the information in these organizations, can enhance the accuracy of information [34]. Additionally, the heuristic-systematic model suggests that the public is inclined to trust the information provided through these sources because of its authority and thereby more motivated to follow the recommendations that these sources offer [35]. By contrast, inexpert sources are those lacking expertise background or professional gatekeepers, namely celebrities, social media influencers, and social contacts that are not doctors. Therefore,

individuals relying on expert versus inexpert sources for information about COVID-19 may demonstrate different patterns of protective behaviors. Given these differences, we predicted that individuals relying on expert sources for information about COVID-19 should be more driven to engage in protective actions whereas reliance on inexpert sources should be related to engagement in protective actions negatively. As mentioned earlier, a repeated cross-sectional investigation will be employed. Hence, an additional question is whether these relationships changed throughout the duration of this study. Three protective actions were assessed: staying at home, wearing a face mask, and washing hands.

Methods

Overview

A six-wave repeated cross-sectional survey was conducted between February 7 and April 23, 2020, in collaboration with a large company that provides sampling services in China. Every other week, an online survey was distributed to a convenience sample of residents in six major cities in China. These cities were Beijing, Shanghai, Guangzhou, and Shenzhen, which are the four largest cities in China, as well as Wuhan, where the first COVID-19 cases were discovered [2], and Hangzhou, another city among the cities with the most reported cases [36].

Our survey started on February 7, 2020. Although cases were first found in Wuhan in late December 2019, the Chinese government did not inform the public that COVID-19 could be transmitted between humans until January 20 [37]. On January 23, Wuhan was locked down [38], which started a series of executive orders on travel bans and wearing face mask [13]. We did not start our research until February 7 because January 24 was the Lunar New Year's Eve, which started a weeklong holiday. Therefore, we could not start our study until early February.

The data collection of wave 1 lasted from February 7-14, 2020. The second wave started on February 20 because most businesses in China restarted by late February and early March [39]. Thus, we wanted to investigate how the resumption of business might have affected our proposed relationships. Given the time difference between these two waves, we decided to collect our data every other week.

The lift of the lockdown in Wuhan on April 7, 2020, signaled the progress of pandemic control [40]. We collected the last wave (April 16-23) of data after April 7 to examine whether the lift of Wuhan's lockdown might have changed our participants' responses.

Sample

Table 1 presents the characteristics of the final sample in each wave. We matched the education and age of our sample to the national population. The most recent national census available to the public shows that around 14% of Chinese people received an associate's degree or higher [41]. We also used this census to calculate the proportion of age strata in our sample: aged 18-30 years (19%), 31-45 years (26%), and 46 years and older (55%). However, this quota of education and age did not always match our sample characteristics in all waves.

Across all waves, there was no significant difference in biological sex (χ^2_5=5.56, P=.35) and city of residence (χ^2_{25}=6.99, P>.99). However, our participants differed significantly between waves in their age ($F_{5,901.48}$=5.75, P<.001; one assumption of one-way variance of analysis is the homogeneity of variances in the dependent variable; however, this assumption was violated when age was compared across waves, so Welch was used to compare differences between waves), education (χ^2_5=27.49, P<.001), and income (χ^2_5=44.88, P<.001).

Table 1. Sample characteristics across waves.

Characteristics	Wave 1 (n=321)	Wave 2 (n=319)	Wave 3 (n=315)	Wave 4 (n=343)	Wave 5 (n=329)	Wave 6 (n=315)
Sex, n (%)						
Male	154 (48)	164 (51.4)	141 (44.8)	157 (45.8)	153 (46.5)	163 (51.7)
Female	167 (52)	155 (48.6)	174 (55.2)	186 (54.2)	176 (53.5)	152 (48.3)
Age (years), n (%)						
18-30	55 (17.1)	64 (20.1)	63 (20)	84 (24.5)	74 (22.5)	60 (19)
31-45	97 (30.2)	82 (25.7)	87 (27.6)	110 (32.1)	86 (26.1)	80 (25.4)
≥46	169 (52.6)	173 (54.2)	165 (52.4)	149 (43.4)	169 (51.4)	175 (55.6)
Education, n (%)						
Middle school or lower	38 (11.8)	21 (6.6)	24 (7.6)	10 (2.9)	23 (7)	50 (15.9)
High school	234 (72.9)	252 (79)	249 (79)	265 (77.3)	250 (76)	222 (70.5)
Associate's degree or higher	49 (15.3)	46 (14.4)	42 (13.3)	68 (19.8)	56 (17)	43 (13.7)
Household monthly income (US $)						
≤500, n (%)	15 (4.7)	9 (2.8)	14 (4.4)	7 (2)	14 (4.3)	25 (7.9)
501-714.29, n (%)	31 (9.7)	29 (9.1)	22 (7)	25 (7.3)	21 (6.4)	36 (11.4)
714.3-1142.86, n (%)	55 (17.1)	41 (12.9)	48 (15.2)	55 (16)	49 (14.9)	75 (23.8)
1142.87-1785.71, n (%)	81 (25.2)	95 (29.8)	87 (27.6)	93 (27.1)	93 (28.3)	89 (28.3)
1785.72-5500, n (%)	113 (35.2)	126 (39.5)	128 (40.6)	138 (40.2)	130 (39.5)	79 (25.1)
5500.01-11,928.57, n (%)	20 (6.2)	9 (2.8)	9 (2.9)	16 (4.7)	16 (4.9)	8 (2.5)
≥11,928.58, n (%)	6 (1.9)	10 (0.31)	7 (2.2)	9 (2.6)	6 (1.8)	3 (1)
Mean (SD)	4.03 (1.32)	4.15 (1.23)	4.1 (1.24)	4.21 (1.19)	4.14 (1.24)	3.63 (1.31)
City of residence, n (%)						
Beijing	55 (17.1)	53 (16.6)	52 (16.5)	59 (17.2)	52 (15.8)	51 (16.2)
Shanghai	54 (16.8)	53 (16.6)	51 (16.2)	58 (16.9)	52 (15.8)	52 (16.5)
Guangzhou	53 (16.5)	53 (16.6)	54 (17.1)	71 (20.7)	54 (16.4)	50 (15.9)
Shenzhen	53 (16.5)	51 (16)	53 (16.8)	51 (14.9)	64 (19.5)	53 (16.8)
Wuhan	52 (16.2)	53 (16.6)	52 (16.5)	50 (14.6)	53 (16.1)	53 (16.8)
Hangzhou	54 (16.8)	56 (17.6)	53 (16.6)	54 (15.7)	54 (16.4)	56 (17.8)

Measures

Table 2 presents the reliability and descriptive statistics of independent and dependent variables in this study. Reliance on expert sources was measured by asking participants to indicate the extent to which their major source of information about COVID-19 was government health departments, government administrations, official media, medical institutes, medical experts, family and friends who are doctors, or the World Health Organization and other health organizations outside China (1=strongly disagree, 7=strongly agree). Reliance on inexpert sources was assessed by the same question except that the sources were replaced with celebrities, social media influencers, family and friends who are not doctors, and other social contacts who are not doctors. The reliability of these two variables at all waves reached .7 or above, except for reliance on inexpert sources, which was .66 at wave 4.

Gore and Bracken's [42] 7-point Likert scale (1=strongly disagree, 7=strongly agree) was adapted to measure perceived severity, perceived susceptibility, response efficacy, and self-efficacy in controlling COVID-19. Specifically, *perceived severity* was measured with three questions ("COVID-19 is a very serious disease/will pose a severe threat to my health/will pose a severe threat to others' safety"), and *perceived susceptibility* was measured with two items ("My chance to get COVID-19 is high" and "I can get COVID-19 from others"). *Response efficacy* was assessed with two items ("modern medical knowledge can control COVID-19" and "COVID-19 can be cured as long as one follows doctors' recommendations"), and *self-efficacy* was assessed with three items ("I can follow the recommended acts to protect myself from COVID-19," "I have no difficulty in performing those protective behaviors that the government recommended," "I can master how to perform recommended actions"). The reliability of these three variables reached .7 or above across all waves except for self-efficacy, which was .69 at wave 2.

Table 2. Cronbach alpha, means, and SDs of major variables.

Variables	Wave 1		Wave 2		Wave 3		Wave 4		Wave 5		Wave 6	
	α	Mean (SD)	α	Mean (SD)	α	Mean (SD)	α	Mean (SD)	α	Mean (SD)	α	Mean (SD)
Expert sources	.78	5.41 (0.85)	.79	5.61 (0.79)	.73	5.61 (0.68)	.77	5.64 (0.74)	.78	5.60 (0.71)	.78	5.76 (0.73)
Inexpert sources	.70	4.00 (1.04)	.75	3.95 (1.04)	.75	3.84 (1.03)	.66	3.93 (0.92)	.80	3.87 (1.04)	.78	3.98 (1.08)
Perceived severity	.80	6.09 (1.03)	.83	6.26 (0.99)	.78	6.10 (1.00)	.81	6.23 (0.99)	.72	6.08 (0.91)	.81	6.24 (0.98)
Perceived susceptibility	.72	4.24 (1.56)	.73	4.30 (1.63)	.79	3.92 (1.59)	.70	4.05 (1.50)	.74	4.06 (1.42)	.83	4.26 (1.59)
Response efficacy	.70	5.42 (1.18)	.80	5.28 (1.33)	.72	5.39 (1.17)	.77	5.35 (1.27)	.70	5.32 (1.14)	.72	5.57 (1.13)
Self-efficacy	.77	5.89 (0.95)	.69	5.86 (0.90)	.71	5.83 (0.87)	.70	5.95 (0.82)	.73	5.84 (0.81)	.78	5.90 (0.91)
Staying at home	N/A[a]	4.10 (1.00)	N/A	4.15 (0.99)	N/A	3.99 (0.96)	N/A	3.76 (1.08)	N/A	3.60 (1.09)	N/A	3.57 (1.11)
Wearing a face mask	N/A	4.75 (0.89)	N/A	4.78 (0.81)	N/A	4.79 (0.78)	N/A	4.81 (0.74)	N/A	4.80 (0.70)	N/A	4.70 (0.86)
Washing hands	N/A	4.72 (0.81)	N/A	4.80 (0.57)	N/A	4.75 (0.64)	N/A	4.70 (0.75)	N/A	4.74 (0.60)	N/A	4.68 (0.67)

[a]N/A: not applicable.

Personal engagement in protective measures was assessed through three 5-point Likert questions. Participants were asked how often they went out during the past 7 days (1=never, 2=once or twice, 3=three or four times, 4=five or six times, 5=seven times or more). We reverse coded participants' response to this question, so the large number indicates *staying at home* more often. We also asked participants how often they wore a face mask and washed their hands during the past 7 days (1=never, 2=rarely, 3=sometimes, 4=often, 5=all the time). Again, larger numbers indicate higher frequency of *wearing a face mask* and *washing hands*.

Control variables were biological sex, age, education (recoded as 1=middle school or lower, 2=high school, 3=associate's degree or higher), household monthly income, city of residence, self-rated health condition (1=very unhealthy, 5=very healthy), and knowledge. *Knowledge* was measured with 17 questions on the transmission of COVID-19, its medication, vulnerable population, and prevention methods. Participants received one point whenever they made a correct option. This made the maximum score 42 points.

Data Analysis

We employed the Kruskal-Wallis H test to examine if there was any difference between engagement in the three protective behaviors and if the level of engagement in these behaviors differed across time. In addition, we conducted repeated ordinal regression through SPSS 25 (IBM Corp) to test our predictions. At each wave, the dependent variables were entered into the model separately, along with control variables and independent variables. This analysis was repeated six times. Log odds ratios (ORs) and ORs along with their 95% CIs were reported to indicate the relationship between two variables.

The ordinal regression results are shown in the tables in the next section. Given the volume of these findings, results were presented separately with different sets of independent variables. Yet, ordinal regression was conducted with all independent variables listed in the tables.

Results

Engagement in Protective Behaviors

Table 3 presents results of the comparisons between engagement in three protective behaviors. Significant differences were found in staying at home across all waves (χ^2_5=110.01, P<.001). However, no significant differences were found in wearing a face mask (χ^2_5=8.07, P=.15) and washing hands (χ^2_5=10.81, P=.06) across time. In addition, across all six waves, we found significant differences consistently in the level of engagement in all three behaviors (Table 3).

Table 3. Differences in engagement in three protective behaviors across time.

Behaviors	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Chi-square (*df*)	*P* value
Staying at home, mean	4.10	4.15	3.99	3.76	3.60	3.57	110.01 (5)	<.001
Wearing a face mask, mean	4.75	4.78	4.79	4.81	4.80	4.70	8.07 (5)	.15
Washing hands, mean	4.72	4.80	4.75	4.70	4.74	4.68	10.81 (5)	.06
Chi-square (*df*)	205.69 (2)	225.46 (2)	252.35 (2)	331.43 (2)	367.66 (2)	344.63 (2)	N/A[a]	N/A
P value	<.001	<.001	<.001	<.001	<.001	<.001	N/A	N/A

[a]N/A: not applicable.

The Effects of Perceived Severity and Perceived Susceptibility

Table 4 presents how perceived severity and perceived susceptibility predicted engagement in the three protective behaviors across time. Perceived severity of COVID-19 predicted staying at home positively at waves 2 and 6 (Table 4). Individuals perceiving COVID-19 as more severe were more likely to wear a face mask at waves 1 and 5 (Table 4). The effect of perceived severity on washing hands was significant at waves 2, 4, 5, and 6 (Table 4). Conversely, perceived susceptibility to COVID-19 only predicted staying at home at waves 1 and 3, and both relationships were negative (Table 4). The effects of perceived susceptibility on wearing a face mask and washing hands were not significant.

Table 4. The effects of perceived severity and perceived susceptibility on engagement in protective behaviors across time.

Time and protective behaviors	Perceived severity		Perceived susceptibility	
	Log OR[a] (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)
Wave 1				
Staying at home	–0.18 (–0.42 to 0.06)	0.84 (0.66 to 1.06)	–0.22 (–0.37 to –0.07)**	0.80 (0.69 to 0.93)
Wearing a face mask	0.46 (0.05 to 0.87)*	1.59 (1.06 to 2.39)	–0.07 (–0.38 to 0.24)	0.93 (0.69 to 1.27)
Washing hands	0.31 (–0.04 to 0.66)	1.36 (0.96 to 1.93)	0.01 (–0.24 to 0.26)	1.01 (0.79 to 1.30)
Wave 2				
Staying at home	0.26 (0.02 to 0.51)*	1.30 (1.02 to 1.66)	–0.06 (–0.20 to 0.09)	0.95 (0.82 to 1.09)
Wearing a face mask	0.34 (–0.09 to 0.77)	1.41 (0.92 to 2.17)	–0.29 (–0.63 to 0.04)	0.75 (0.53 to 1.04)
Washing hands	0.53 (0.20 to 0.85)**	1.69 (1.22 to 2.35)	–0.07 (–0.31 to 0.17)	0.93 (0.73 to 1.19)
Wave 3				
Staying at home	–0.12 (–0.36 to 0.11)	0.89 (0.70 to 1.12)	–0.21 (–0.35 to –0.06)**	0.81 (0.70 to 0.94)
Wearing a face mask	0.05 (–0.36 to 0.47)	1.06 (0.70 to 1.60)	–0.15 (–0.45 to 0.14)	0.86 (0.64 to 1.16)
Washing hands	–0.06 (–0.40 to 0.27)	0.94 (0.67 to 1.31)	–0.16 (–0.37 to 0.05)	0.85 (0.69 to 1.05)
Wave 4				
Staying at home	–0.16 (–0.38 to 0.06)	0.85 (0.68 to 1.06)	–0.06 (–0.20 to 0.09)	0.94 (0.82 to 1.09)
Wearing a face mask	0.04 (–0.37 to 0.45)	1.04 (0.69 to 1.57)	–0.02 (–0.34 to 0.31)	0.99 (0.72 to 1.36)
Washing hands	0.37 (0.10 to 0.64)**	1.45 (1.10 to 1.90)	–0.15 (–0.38 to 0.07)	0.86 (0.69 to 1.07)
Wave 5				
Staying at home	0.11 (–0.14 to 0.36)	1.12 (0.87 to 1.43)	0.01 (–0.15 to 0.16)	1.01 (0.86 to 1.18)
Wearing a face mask	0.55 (0.12 to 0.98)*	1.73 (1.13 to 2.66)	0.12 (–0.19 to 0.43]	1.13 (0.83 to 1.54)
Washing hands	0.41 (0.07 to 0.74)*	1.50 (1.07 to 2.10)	–0.08 (–0.31 to 0.15)	0.92 (0.73 to 1.16)
Wave 6				
Staying at home	0.28 (0.04 to 0.51)*	1.32 (1.04 to 1.66)	0.002 (–0.14 to 0.14)	1.00 (0.87 to 1.15)
Wearing a face mask	0.26 (–0.05 to 0.57)	1.30 (0.95 to 1.77)	–0.06 (–0.31 to 0.18)	0.94 (0.74 to 1.20)
Washing hands	0.31 (0.03 to 0.59)*	1.36 (1.03 to 1.81)	–0.09 (–0.40 to 0.23)	0.92 (0.67 to 1.25)

[a]OR: odds ratio.
*P<.05.
**P<.01.

The Effects of Response Efficacy and Self-Efficacy

Table 5 shows how response efficacy and self-efficacy affected engagement in protective actions across time. At wave 1, response efficacy predicted staying at home and washing hands positively (Table 5). After wave 1, its effect on protective behaviors became weak. Individuals who reported higher levels of response efficacy were more likely to stay at home at wave 2 and wash hands at wave 4 (Table 5). Response efficacy was not significantly associated with wearing a face mask at any time. Self-efficacy did not predict any protective behavior at any time.

Table 5. The effects of response efficacy and self-efficacy on engagement in protective behaviors across time.

Time and protective behaviors	Response efficacy		Self-efficacy	
	Log OR[a] (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)
Wave 1				
Staying at home	0.30 (0.08 to 0.52)**	1.35 (1.08 to 1.68)	–0.04 (–0.33 to 0.26)	0.97 (0.72 to 1.30)
Wearing a face mask	–0.01 (–0.42 to 0.41)	0.93 (0.69 to 1.27)	0.24 (–0.27 to 0.74)	1.27 (0.77 to 2.10)
Washing hands	0.36 (0.04 to 0.68)*	1.43 (1.04 to 1.97)	–0.10 (–0.53 to 0.34)	0.91 (0.59 to 1.40)
Wave 2				
Staying at home	0.31 (0.11 to 0.50)**	1.36 (1.11 to 1.66)	–0.04 (–0.34 to 0.27)	0.96 (0.71 to 1.31)
Wearing a face mask	0.05 (–0.36 to 0.46)	1.05 (0.70 to 1.59)	0.24 (–0.33 to 0.81)	1.27 (0.73 to 2.24)
Washing hands	0.01 (–0.30 to 0.32)	1.01 (0.74 to 1.38)	0.37 (–0.10 to 0.83)	1.45 (0.91 to 2.30)
Wave 3				
Staying at home	0.08 (–0.13 to 0.28)	1.08 (0.88 to 1.32)	–0.15 (–0.44 to 0.14)	0.86 (0.64 to 1.15)
Wearing a face mask	–0.07 (–0.48 to 0.35)	0.94 (0.62 to 1.42)	–0.21 (–0.76 to 0.34)	0.81 (0.47 to 1.40)
Washing hands	–0.14 (–0.43 to 0.15)	0.87 (0.65 to 1.17)	0.25 (–0.15 to 0.65)	1.28 (0.86 to 1.91)
Wave 4				
Staying at home	–0.09 (–0.27 to 0.08)	0.91 (0.77 to 1.09)	0.02 (–0.28 to 0.31)	1.02 (0.76 to 1.37)
Wearing a face mask	0.25 (–0.09 to 0.58)	1.28 (0.92 to 1.79)	0.03 (–0.52 to 0.58)	1.04 (0.60 to 1.79)
Washing hands	0.26 (0.01 to 0.51)*	1.30 (1.01 to 1.66)	–0.07 (–0.47 to 0.33)	0.94 (0.63 to 1.39)
Wave 5				
Staying at home	0.18 (–0.04 to 0.39)	1.19 (0.96 to 1.48)	–0.17 (–0.47 to 0.13)	0.84 (0.63 to 1.14)
Wearing a face mask	0.15 (–0.27 to 0.58)	1.17 (0.76 to 1.78)	–0.15 (–0.70 to 0.40)	0.86 (0.50 to 1.49)
Washing hands	0.02 (–0.29 to 0.33)	1.02 (0.75 to 1.39)	0.28 (–0.13 to 0.69)	1.33 (0.88 to 2.00)
Wave 6				
Staying at home	0.03 (–0.20 to 0.25)	1.03 (0.82 to 1.28)	0.05 (–0.26 to 0.36)	1.05 (0.77 to 1.43)
Wearing a face mask	–0.13 (–0.52 to 0.27)	0.88 (0.59 to 1.31)	0.15 (–0.32 to 0.62)	1.16 (0.73 to 1.86)
Washing hands	–0.09 (–0.40 to 0.23)	0.92 (0.67 to 1.25)	0.19 (–0.19 to 0.58)	1.21 (0.83 to 1.78)

[a]OR: odds ratio.
*P<.05.
**P<.01.

The Effects of Reliance on Expert Versus Inexpert Sources

Table 6 demonstrates how individuals' reliance on expert versus inexpert sources for information about COVID-19 might affect their engagement in the three protective actions across time. Reliance on expert sources did not predict engagement in any protective behaviors at wave 1, and only predicted wearing a face mask at wave 2 (Table 6). Starting from wave 3, the facilitating effect of expert sources became more prominent. Specifically, reliance on expert sources predicted staying at home positively at waves 3 and 4 (Table 6). In addition to wave 2, individuals relying on expert sources for information about COVID-19 were more likely to wear a face mask at waves 4, 5, and 6 (Table 6). The relationship between reliance on expert sources and washing hands was significant at waves 3, 4, 5, and 6 (Table 6).

The effect of reliance on inexpert sources on protective behaviors was more limited. Reliance on inexpert sources exhibited a *negative* effect on *staying at home* at wave 2 (Table 6). Individuals relying on inexpert sources were *less* likely to *wear a face mask* at wave 2 and *wash hands* at wave 5 (Table 6).

Table 6. The effects of reliance on expert versus inexpert sources on engagement in protective behaviors across time.

Time and protective behaviors	Expert sources		Inexpert sources	
	Log OR[a] (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)
Wave 1				
Staying at home	0.06 (–0.25 to 0.37)	1.06 (0.78 to 1.45)	0.04 (–0.18 to 0.25)	1.04 (0.84 to 1.28)
Wearing a face mask	–0.06 (–0.71 to 0.60)	0.94 (0.49 to 1.81)	–0.22 (–0.66 to 0.22)	0.80 (0.52 to 1.24)
Washing hands	0.33 (–0.16 to 0.82)	1.39 (0.85 to 2.28)	–0.34 (–0.72 to 0.05)	0.71 (0.49 to 1.05)
Wave 2				
Staying at home	0.34 (–0.01 to 0.69)	1.40 (0.99 to 1.99)	–0.25 (–0.49 to –0.02)*	0.78 (0.62 to 0.99)
Wearing a face mask	0.68 (0.01 to 1.34)*	1.97 (1.01 to 3.83)	–0.50 (–0.97 to –0.03)*	0.61 (0.38 to 0.97)
Washing hands	0.39 (–0.13 to 0.92)	1.48 (0.87 to 2.51)	–0.27 (–0.65 to 0.12)	0.77 (0.52 to 1.12)
Wave 3				
Staying at home	0.50 (0.11 to 0.89)*	1.64 (1.11 to 2.42)	–0.20 (–0.43 to 0.03)	0.83 (0.65 to 1.03)
Wearing a face mask	0.64 (–0.12 to 1.40)	1.89 (0.89 to 4.05)	–0.06 (–0.48 to 0.36)	0.94 (0.62 to 1.43)
Washing hands	0.76 (0.21 to 1.30)**	2.13 (1.23 to 3.68)	–0.15 (–0.48 to 0.17)	0.86 (0.62 to 1.19)
Wave 4				
Staying at home	0.49 (0.17 to 0.82)**	1.64 (1.18 to 2.27)	–0.10 (–0.34 to 0.14)	0.91 (0.71 to 1.15)
Wearing a face mask	0.82 (0.20 to 1.44)*	2.26 (1.22 to 4.22)	–0.34 (–0.89 to 0.20)	0.71 (0.41 to 1.23)
Washing hands	0.96 (0.50 to 1.42)***	2.61 (1.65 to 4.12)	–0.04 (–0.39 to 0.31)	0.96 (0.68 to 1.36)
Wave 5				
Staying at home	0.01 (–0.34 to 0.35)	1.01 (0.71 to 1.42)	–0.08 (–0.30 to 0.14)	0.92 (0.74 to 1.15)
Wearing a face mask	0.64 (0.09 to 1.20)*	1.90 (1.09 to 3.30)	–0.35 (–0.80 to 0.11)	0.71 (0.45 to 1.11)
Washing hands	0.59 (0.13 to 1.05)*	1.80 (1.14 to 2.85)	–0.37 (–0.71 to –0.03)*	0.69 (0.49 to 0.97)
Wave 6				
Staying at home	0.08 (–0.30 to 0.46)	1.08 (0.74 to 1.58)	–0.22 (–0.44 to 0.003)	0.80 (0.64 to 1.00)
Wearing a face mask	0.64 (0.09 to 1.19)*	1.90 (1.10 to 3.28)	–0.19 (–0.54 to 0.16)	0.83 (0.58 to 1.17)
Washing hands	0.74 (0.27 to 1.22)**	2.10 (1.31 to 3.39)	–0.05 (–0.34 to 0.24)	0.95 (0.71 to 1.28)

[a]OR: odds ratio.
*P<.05.
**P<.01.
***P<.001.

The Effects of Control Variables

Knowledge did not predict *staying at home* at any time. Individuals equipped with more knowledge were more likely to *wear a face mask* at wave 3 (OR 1.14, 95% CI 1.05-1.25; P<.01). The relationship between knowledge and *washing hands* was only significant and positive at wave 3 (OR 1.09, 95% CI 1.02-1.17; P<.05) and wave 6 (OR 1.08, 95% CI 1.01-1.15; P<.05).

The self-rated health condition predicted *wearing a face mask* (OR 2.21, 95% CI 1.27-3.85; P<.01) and *washing hands* positively at wave 4 (OR 1.63, 95% CI 1.10-2.41; P<.05). At wave 6, the relationship between self-rated health condition and *staying at home* was positive (OR 1.43, 95% CI 1.02-2.00; P<.05).

Income predicted *staying at home negatively* at wave 3 (OR 0.77, 95% CI 0.64-0.93; P<.01), wave 4 (OR 0.81, 95% CI 0.67-0.97; P<.05), and wave 6 (OR 0.83, 95% CI 0.69-1.00; P<.01). Individuals with a greater household monthly income were *more* likely to *wear a face mask* at wave 1 (OR 1.67, 95% CI 1.19-2.35; P<.01), wave 2 (OR 1.49, 95% CI 1.02-2.17; P<.05), wave 4 (OR 1.47, 95% CI 1.00-2.16; P<.05), and wave 6 (OR 1.43, 95% CI 1.08-1.89; P<.05). The relationship between income and *washing hands* was *positive* at wave 1 (OR 1.37, 95% CI 1.04-1.79; P<.05) and wave 4 (OR 1.37, 95% CI 1.06-1.77; P<.05).

Compared to women, men *washed hands* less often at wave 4 (OR 0.38, 95% CI 0.20-0.72; P<.01), wave 5 (OR 0.43, 95% CI 0.23-0.80; P<.01), and wave 6 (OR 0.50, 95% CI 0.28-0.92; P<.05). *Age* predicted *washing hands* positively at wave 1 (OR 1.04, 95% CI 1.01-1.08; P<.05), wave 4 (OR 1.06, 95% CI 1.02-1.09; P<.01), and wave 6 (OR 1.04, 95% CI 1.00-1.07; P<.05). At wave 4, participants with a high school degree

washed hands less often than those with an associate's degree or above (OR 0.36, 95% CI 0.15-0.89; *P*<.05).

When it comes to city differences, residents in Wuhan, where COVID-19 cases were first discovered, were used as the reference group. No significant difference was found in *wearing a face mask* and *washing hands* across all waves, except that residents in Beijing reported to *wear a face mask* more often than those in Wuhan (OR 9.69, 95% CI 1.09-86.38; *P*<.05). However, residents in Wuhan *stayed at home* more often than those in the other cities at most times, as Tables 7 and 8 shows.

Table 7. City differences in staying at home waves 1, 2, and 3 (Wuhan was used as the reference group).

City	Wave 1		Wave 2		Wave 3	
	Log OR[a] (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)
Beijing	–0.62 (–1.38 to 0.13)	0.54 (0.25 to 1.14)	–1.76 (–2.57 to –0.95)***	0.17 (0.08 to 0.39)	–0.84 (–1.60 to –0.07)*	0.43 (0.20 to 0.93)
Shanghai	–0.99 (–1.75 to –0.22)*	0.37 (0.17 to 0.80)	–1.74 (–2.56 to –0.92)***	0.18 (0.08 to 0.40)	–1.32 (–2.10 to –0.54)**	0.27 (0.12 to 0.58)
Guangzhou	–0.87 (–1.62 to –0.11)*	0.42 (0.20 to 0.89)	–1.75 (–2.58 to –0.92)***	0.17 (0.08 to 0.40)	–0.72 (–1.49 to 0.04)	0.49 (0.23 to 1.04)
Shenzhen	–01.15 (–1.93 to –0.36)**	0.32 (0.15 to 0.70)	–1.58 (–2.41 to –0.75)***	0.21 (0.09 to 0.47)	–0.81 (–1.58 to –0.04)*	0.44 (0.21 to 0.96)
Hangzhou	–1.01 (–1.77 to –0.25)**	0.37 (0.17 to 0.78)	–1.33 (–2.18 to 0.47)**	0.27 (0.11 to 0.62)	–1.18 (–1.95 to –0.42)**	0.31 (0.14 to 0.66)

[a]OR: odds ratio.
*P<.05.
**P<.01.
***P<.001.

Table 8. City differences in staying at home waves 4, 5, and 6 (Wuhan was used as the reference group).

City	Wave 4		Wave 5		Wave 6	
	Log OR[a] (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)	Log OR (95% CI)	OR (95% CI)
Beijing	–1.26 (–2.02 to –0.51)**	0.28 (0.13 to 0.60)	–0.14 (–0.90 to 0.61)	0.87 (0.41 to 1.84)	–0.49 (–1.26 to 0.27)	0.61 (0.28 to 1.31)
Shanghai	–1.35 (–2.11 to –0.58)**	0.26 (0.12 to 0.56)	–0.63 (–1.40 to 0.14)	0.53 (0.25 to 1.15)	–0.99 (–1.78 to –0.21)*	0.37 (0.17 to 0.81)
Guangzhou	–1.35 (–2.07 to –0.64)***	0.26 (0.13 to 0.53)	–1.28 (–2.04 to –0.52)**	0.28 (0.13 to 0.59)	–0.52 (–1.32 to 0.28)	0.59 (0.27 to 1.32)
Shenzhen	–1.25 (–2.02 to –0.48)**	0.29 (0.13 to 0.62)	–0.73 (–1.44 to –0.02)*	0.48 (0.24 to 0.99)	–0.44 (–1.23 to 0.35)	0.65 (0.29 to 1.42)
Hangzhou	–1.75 (–2.52 to –0.98)***	0.17 (0.08 to 0.38)	–1.02 (–1.75 to –0.30)**	0.36 (0.17 to 0.74)	–0.55 (–1.31 to 0.21)	0.58 (0.27 to 1.23)

[a]OR: odds ratio.
*P<.05.
**P<.01.
***P<.001.

Summary

Despite inconsistencies, some patterns still emerged. First, reliance on expert sources encouraged protective behaviors, but this effect did not emerge until wave 3 and was stronger on wearing a face mask and washing hands. Second, the discouraging effect of reliance on inexpert sources was limited to wave 2 except that it predicted washing hands negatively at wave 5. In addition, perceived severity exhibited a stronger effect on protective behaviors than perceived susceptibility. Furthermore, self-efficacy was not associated with engaging in protective behaviors, whereas the effect of response efficacy was limited to waves 1 and 2. Among all control variables, the effect of knowledge was limited, whereas the city of residence exhibited a stronger effect on staying at home.

Discussion

Principal Findings

The COVID-19 pandemic triggered research on what factors affected individuals' engagement in protective behaviors [7-12]. This study is built upon EPPM, a theoretical framework that explains how risk perception and efficacy appraisal might affect individuals' engagement in protective behaviors [13]. In addition, given the volume of misinformation about preventive measures against COVID-19 [33], we extended EPPM and the

extant research on protective actions against COVID-19 by recognizing the value of accurate information and considering Chinese individuals' reliance on expert versus inexpert information sources. Further, differences across time and between three target behaviors were also revealed. The patterns of our findings previously summarized provide important implications on health education and suggest the intertwined relationship between one's health behavior and the sociocultural system where these individuals reside.

First, we found that perceived severity could encourage protective behaviors, but their effects were not consistent and different depending on the specific behavior. Taken as a whole, perceived severity predicted washing hands positively at waves 2, 4, 5, and 6, more consistently than wearing a face mask (waves 1 and 5) and staying at home (waves 2 and 6). The inconsistency might be related to the executive orders that the Chinese government issued, which forced individuals to wear a face mask in public and placed them in quarantine [13]. Therefore, in this study, wearing a face mask and staying at home were not entirely autonomous decisions but more because of compliance with the executive orders. However, washing hands was not required, and it was impossible to ensure that everyone washed their hands as recommended. Thus, how often individuals washed their hands was likely derived from their evaluation of the risk.

Surprisingly, perceived susceptibility predicted staying at home negatively at waves 1 and 3. The post hoc analysis found that at both waves the common predictor of perceived susceptibility was self-rated health condition (wave 1: OR 0.72, 95% CI 0.56-0.92; $P<.05$; wave 3: OR 0.65, 95% CI 0.50-0.86; $P<.01$), and older participants reported a worse health condition (wave 1: OR 0.97, 95% CI 0.95-0.99; $P<.01$; wave 3: OR 0.95, 95% CI 0.93-0.97; $P<.001$). Therefore, among older participants, there might be a gap between risk perception and behavior. Although they realized that they could be subject to COVID-19, they still went out. This suggests that health education for seniors should focus on bridging the perception-behavior gap.

Overall, the effect of perceived susceptibility on protective behaviors was minimal. However, the impact of perceived susceptibility should not be dismissed. For example, protection motivation theory contends that human behavior is a function of the perceived severity of the threat, perceived susceptibility to the threat, and response efficacy, and no behavior is performed if any of these predictors are zero [14]. Although more empirical evidence is needed to understand whether health education in China during the pandemic lacks information on susceptibility, this result suggested that subsequent education should highlight the chance that certain populations are vulnerable to the pandemic.

In addition to risk perceptions, our results showed that response efficacy only predicted staying at home at waves 1 and 2, and washing hands at wave 1. Hence, at the early stage of the outbreak, individuals engaged in preventive measures because perhaps they believed these actions were effective to protect them against the given threat. This suggests that practitioners may want to adjust the emphasis of health education as time passes. Specifically, elevating response efficacy of the target audience may be important at the early stage of the outbreak.

By contrast, self-efficacy did not predict any protective behavior at any time. One possible reason is that our measure of self-efficacy addressed overall confidence in performing preventive measures instead of specific preventive actions. However, there might be differences in the level of difficulty in performing these three protective behaviors. Thus, our measure might not have assessed this subtle difference.

It is important to note that EPPM research tends to test the aggregate effects of perceived severity and perceived susceptibility as well as response efficacy and self-efficacy on protective behaviors [18,21,22,42,43]. However, we demonstrated the separate effects of these variables, and we found their distinct effects. This suggests that perceived severity versus perceived susceptibility (response efficacy vs self-efficacy) may be essentially different, which needs further study.

In addition to testing EPPM, our results demonstrated how reliance on expert versus inexpert sources might affect Chinese individuals' engagement in protective actions. Our findings reveal that the positive effect of expert sources did not emerge until wave 3 when most businesses restarted [39]. The post hoc analysis found that, controlling for knowledge, self-rated health condition, and demographic variables, reliance on expert sources at wave 1 was significantly lower than all other waves (wave 2: OR 2.26, 95% CI 1.72-2.98; $P<.001$; wave 3: OR 1.56, 95% CI 1.19-2.04; $P<.01$; wave 4: OR 1.75, 95% CI 1.34-2.29; $P<.001$; wave 5: OR 1.54, 95% CI 1.17-2.02; $P<.01$; wave 6: OR 1.57, 95% CI 1.20-2.06; $P<.01$). One explanation is that it took time for the Chinese public to develop trust in these expert sources and follow the messages that these sources delivered. Expert sources in China, such as official media and health departments, are under strict control by the Chinese government, which was blamed for their failure to provide timely responses to COVID-19 during its early outbreak. This might have affected Chinese individuals' trust in these expert sources given their close connection with the government. However, the aggressive actions that the government took controlled the spread of the pandemic and made the number of cases start to decline in late February 2020 [3,13]. Therefore, at wave 3, which started in early March, Chinese individuals might have gained more trust in these expert sources, making them more willing to comply with the recommendations that these sources offered. This suggests that individuals' trust in information sources may exhibit a critical impact on their health behavior. Furthermore, this finding suggests that the conventional approach to persuading the public to engage in protective behaviors during the pandemic, which centers on knowledge provision, may not be effective. A more important mission might be to help the public develop trust in the community of public health practitioners including those working for the government. Therefore, a perspective of public relations is needed in future research and practices on health education.

In contrast, reliance on inexpert sources did not affect protective behaviors most of the time, except that these sources discouraged preventive measures at wave 2. This shows that

our participants might have realized the risks of inexpert sources in information provision, so they did not follow this information. Although these findings are promising, information literacy should still be a focus of future health education and campaigns, especially those vulnerable to health misinformation, such as seniors and less educated individuals.

Additionally, the significant effect of reliance on inexpert sources was limited to wave 2. One possible explanation is that the public interest changed as time passed. In January and February 2020, the public may have been concerned about how to control and treat COVID-19. However, the restart of businesses might have signaled that the pandemic was under control. By then, individuals may have been more concerned about economic recession and recovery. Hence, after wave 2, the focus of the information exchanged between inexpert sources might have changed, which made reliance on these sources not significantly related to taking preventive measures.

Finally, the effects of several control variables warrant discussion. The impact of knowledge on protective behaviors was limited, and residents in Wuhan stayed at home more than participants in other cities at most times. These two findings can be explained by the influence of executive orders. The lockdown of Wuhan lasted more than 2 months, so naturally, participants from Wuhan stayed at home more. Additionally, the limited influence of knowledge suggests that Chinese individuals' engagement in protective behaviors might not be a result of their autonomous decisions but compliance with executive orders. Although this approach to behavior change controlled the spread of COVID-19 in China [3,13], the duration of its effect is questionable, which future research needs to investigate.

Limitations and Future Research

These findings must be interpreted with several caveats. First, the cross-sectional nature of this study makes it impossible to build causal relationships between variables. Second, our study uses self-reported data. This method relies on participants' memory and can be subject to social desirability.

In addition, as previously explained, Chinese individuals performed these protective actions partly because of their compliance with strict law enforcement and executive orders issued by China's government. This might explain why our participants' responses to questions measuring their engagement in protective behaviors were skewed. Furthermore, this might affect the validity of responses that our participants provided. Hence, social desirability must be considered when results are interpreted.

Although we matched the age and the education level of our sample to the national population in China, the generalizability of our sample may still be a limitation. Moreover, the proportions of education and age did not match the national population at all waves. The significant differences in education, income, and age between waves might have introduced additional variances and affected the validity of our results.

This study was conducted in China during the COVID-19 pandemic. This particular timing and geographic location might limit the generalizability of our results. Cross-cultural comparisons and longitudinal observations can be valuable directions for future research.

Our measures of self-efficacy and knowledge could also affect the validity of our findings. As mentioned earlier, the measure of self-efficacy did not specify the preventive behavior. Moreover, we self-created our scale of knowledge based on relevant information from the media. Established measures based on a manual provided by health departments would be more valid.

It is important to note that our definition of risk perception was limited to cognitive appraisal, which may dismiss the effect of affective responses. Future inquiries are needed to understand how cognitive and affective appraisals of risks may affect individuals' engagement in protective behaviors during the pandemic.

Finally, as argued earlier, whether Chinese individuals engaged in protective behavior might partly be a result of strict executive orders. Thus, Chinese individuals' attitude toward the political system may play a part in their engagement in protective behaviors against COVID-19. This implication may also apply in other countries such as the United States, where pandemic control has been politicalized [44,45]. Therefore, future research may need to examine how variables such as political interest and political orientation may affect one's health behavior.

Conclusion

This study provides empirical evidence on what affected Chinese individuals' engagement in protective behaviors against COVID-19 between February and April 2020. Given the authoritarian political system in the media, Chinese individuals' engagement in protective behavior might not be an entirely autonomous decision but a result of compliance with executive orders. Our findings demonstrate that expert sources did not encourage protective behaviors until the early stage passed, suggesting that it might take time to develop trust in expert sources. Therefore, the effect of health education may depend on information as well as the relationship between practitioners and the public. This suggests that a perspective of public relations should be considered in future research. In addition, perceived severity could motivate some protective measures, but its effect differed depending on the specific behavior. Furthermore, the facilitating effect of perceived severity lasted throughout the duration of the pandemic but that of response efficacy was limited to the early stage. Hence, practitioners may want to adjust the emphasis of health campaigns depending on the stage of the pandemic.

Acknowledgments

This work was supported by the Department of Education in Guangdong Province under grant 2019SFKC06.

Conflicts of Interest

None declared.

References

1. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). Johns Hopkins University. URL: https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 [accessed 2020-07-31]
2. WHO timeline - COVID-19. World Health Organization. URL: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 [accessed 2020-07-21]
3. Kupferschmidt K, Cohen J. China's aggressive measures have slowed the coronavirus. They may not work in other countries. Science. 2020 Mar 02. URL: https://www.sciencemag.org/news/2020/03/china-s-aggressive-measures-have-slowed-coronavirus-they-may-not-work-other-countries [accessed 2020-07-21]
4. Callaway E, Ledford H, Mallapaty S. Six months of coronavirus: the mysteries scientists are still racing to solve. Nature. 2020 Jul 03. URL: https://www.nature.com/articles/d41586-020-01989-z [accessed 2020-07-21]
5. Community-based health care, including outreach and campaigns, in the context of the COVID-19 pandemic. World Health Organization. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Comm_health_care-2020.1 [accessed 2020-07-21]
6. Collecting open government approaches to COVID-19. Open Government Partnership. URL: https://www.opengovpartnership.org/collecting-open-government-approaches-to-covid-19/ [accessed 2020-07-21]
7. Clements J. Knowledge and behaviors toward COVID-19 among US residents during the early days of the pandemic: cross-sectional online questionnaire. JMIR Public Health Surveill 2020 May 08;6(2):e19161 [FREE Full text] [doi: 10.2196/19161] [Medline: 32369759]
8. Harper C, Satchell L, Fido D, Latzman R. Functional fear predicts public health compliance in the COVID-19 pandemic. Int J Ment Health Addict 2020 Apr 27:1-14 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
9. Li S, Feng B, Liao W, Pan W. Internet use, risk awareness, and demographic characteristics associated with engagement in preventive behaviors and testing: cross-sectional survey on COVID-19 in the United States. J Med Internet Res 2020 Jun 16;22(6):e19782 [FREE Full text] [doi: 10.2196/19782] [Medline: 32501801]
10. Taghrir M, Borazjani R, Shiraly R. COVID-19 and Iranian medical students; a survey on their related-knowledge, preventive behaviors and risk perception. Arch Iran Med 2020 Apr 01;23(4):249-254. [doi: 10.34172/aim.2020.06] [Medline: 32271598]
11. Wise T, Zbozinek T, Michelini G, Hagan CC, Mobbs D. Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States. PsyArXiv Preprint posted online March 19, 2020. [doi: 10.31234/osf.io/dz428]
12. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
13. Lhatoo Y. Never mind China, look to the US for the next big coronavirus crisis. South China Morning Post. 2020 Mar 07. URL: https://www.scmp.com/news/hong-kong/health-environment/article/3074087/never-mind-china-look-us-next-big-coronavirus [accessed 2020-07-21]
14. Rogers R. A protection motivation theory of fear appeals and attitude change. J Psychol 1975 Sep;91(1):93-114. [doi: 10.1080/00223980.1975.9915803] [Medline: 28136248]
15. Janz N, Becker M. The Health Belief Model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
16. Rimal R, Real K. Perceived risk and efficacy beliefs as motivators of change: use of the risk perception attitude (RPA) framework to understand health behaviors. Human Comm Res 2003 Jul 01;29(3):370-399. [doi: 10.1111/j.1468-2958.2003.tb00844.x]
17. Witte K. Putting the fear back into fear appeals: the extended parallel process model. Commun Monogr 1992 Dec;59(4):329-349. [doi: 10.1080/03637759209376276]
18. Witte K. Fear as motivator, fear as inhibitor: using the extended parallel process model to explain fear appeal successes and failures. In: Andersen PA, Guerrero LK, editors. Handbook of Communication and Emotion: Research, Theory, Applications, and Contexts. Cambridge, MA: Academic Press; 1996:423-450.
19. van der Pligt J. Risk perception and self-protective behavior. Eur Psychol 1996 Jan;1(1):34-43. [doi: 10.1027/1016-9040.1.1.34]
20. Rimal RN. Perceived risk and self-efficacy as motivators: understanding individuals' long-term use of health information. J Commun 2001 Dec;51(4):633-654. [doi: 10.1111/j.1460-2466.2001.tb02900.x]
21. Wong N, Cappella J. Antismoking threat and efficacy appeals: effects on smoking cessation intentions for smokers with low and high readiness to quit. J Appl Commun Res 2009;37(1):1-20 [FREE Full text] [doi: 10.1080/00909880802593928] [Medline: 20046966]

22. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000 Oct;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
23. Barkley-Levenson E, Van Leijenhorst L, Galván A. Behavioral and neural correlates of loss aversion and risk avoidance in adolescents and adults. Dev Cogn Neurosci 2013 Jan;3:72-83 [FREE Full text] [doi: 10.1016/j.dcn.2012.09.007] [Medline: 23245222]
24. Hadlaczky G, Hökby S, Mkrtchian A, Wasserman D, Balazs J, Machín N, et al. Decision-making in suicidal behavior: the protective role of loss aversion. Front Psychiatry 2018;9:116. [doi: 10.3389/fpsyt.2018.00116] [Medline: 29674980]
25. de Hoog N, Stroebe W, de Wit JBF. The impact of vulnerability to and severity of a health risk on processing and acceptance of fear-arousing communications: a meta-analysis. Rev Gen Psychol 2007 Sep 01;11(3):258-285. [doi: 10.1037/1089-2680.11.3.258]
26. Floyd DL, Prentice-Dunn S, Rogers RW. A meta-analysis of research on protection motivation theory. J Appl Soc Psychol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
27. Tannenbaum M, Hepler J, Zimmerman R, Saul L, Jacobs S, Wilson K, et al. Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. Psychol Bull 2015 Nov;141(6):1178-1204 [FREE Full text] [doi: 10.1037/a0039729] [Medline: 26501228]
28. Sheeran P, Harris P, Epton T. Does heightening risk appraisals change people's intentions and behavior? A meta-analysis of experimental studies. Psychol Bull 2014 Mar;140(2):511-543. [doi: 10.1037/a0033065] [Medline: 23731175]
29. Lam S. Predicting intention to save water: theory of planned behavior, response efficacy, vulnerability, and perceived efficiency of alternative solutions. J Appl Social Pyschol 2006 Nov;36(11):2803-2824. [doi: 10.1111/j.0021-9029.2006.00129.x]
30. Lewis I, Watson B, White K. Response efficacy: the key to minimizing rejection and maximizing acceptance of emotion-based anti-speeding messages. Accid Anal Prev 2010 Mar;42(2):459-467. [doi: 10.1016/j.aap.2009.09.008] [Medline: 20159067]
31. Nabi R, Prestin A. Unrealistic hope and unnecessary fear: exploring how sensationalistic news stories influence health behavior motivation. Health Commun 2016 Sep;31(9):1115-1126. [doi: 10.1080/10410236.2015.1045237] [Medline: 26886401]
32. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
33. Cuan-Baltazar J, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
34. Shoemaker P, Vos T. Gatekeeping Theory. New York, NY: Routledge; 2009.
35. Cummings L. The "trust" heuristic: arguments from authority in public health. Health Commun 2014;29(10):1043-1056. [doi: 10.1080/10410236.2013.831685] [Medline: 24447008]
36. Diao M, Zhang S, Chen D, Hu W. The novel coronavirus (COVID-19) infection in Hangzhou: an experience to share. Infect Control Hosp Epidemiol 2020 Jul;41(7):874-875 [FREE Full text] [doi: 10.1017/ice.2020.62] [Medline: 32131914]
37. Kuo L. China confirms human-to-human transmission of coronavirus. The Guardian. 2020 Jan 21. URL: https://www.theguardian.com/world/2020/jan/20/coronavirus-spreads-to-beijing-as-china-confirms-new-cases [accessed 2020-07-21]
38. Crossley G. Wuhan lockdown 'unprecedented', shows commitment to contain virus: WHO representative in China. Reuters. 2020 Jan 23. URL: https://www.reuters.com/article/us-china-health-who-idUSKBN1ZM1G9 [accessed 2020-07-21]
39. Singh S, Leng C, Sun Y. China businesses to restart at larger scale by late Feb, early March: Ministry of Transport. Reuters. 2020 Feb 22. URL: https://www.reuters.com/article/us-china-health-business/china-businesses-to-restart-at-larger-scale-by-late-feb-early-march-ministry-of-transport-idUSKCN20G09N [accessed 2020-07-21]
40. Jeffery A. Coronavirus: photos of Wuhan after 11-week lockdown. CNBC. 2020 Apr 08. URL: https://www.cnbc.com/2020/04/08/wuhan-lifts-travel-restrictions-after-11-week-lockdown-see-photos.html [accessed 2020-07-21]
41. The 2018 Yearbook of Statistics in China. National Bureau of Statistics in China. URL: http://www.stats.gov.cn/tjsj/ndsj/2018/indexch.htm [accessed 2020-07-21]
42. Gore T, Bracken C. Testing the theoretical design of a health risk message: reexamining the major tenets of the extended parallel process model. Health Educ Behav 2005 Feb;32(1):27-41. [doi: 10.1177/1090198104266901] [Medline: 15642752]
43. Chen L, Yang X. Using EPPM to evaluate the effectiveness of fear appeal messages across different media outlets to increase the intention of breast self-examination among Chinese women. Health Commun 2019 Oct;34(11):1369-1376. [doi: 10.1080/10410236.2018.1493416] [Medline: 30080982]
44. Painter M, Qiu T. Political beliefs affect compliance with COVID-19 social distancing orders. SSRN J 2020:1. [doi: 10.2139/ssrn.3569098]
45. Calvillo D, Ross B, Garcia R, Smelter T, Rutchick A. Political ideology predicts perceptions of the threat of COVID-19 (and susceptibility to fake news about it). Soc Psychological Pers Sci 2020 Jul 22;11(8):1119-1128. [doi: 10.1177/1948550620940539]

Abbreviations

EPPM: extended parallel process model
OR: odds ratio

Edited by A Kushniruk; submitted 05.08.20; peer-reviewed by D Chrimes, S Kim; comments to author 11.10.20; revised version received 29.10.20; accepted 24.11.20; published 12.01.21.

Please cite as:
Rui JR, Yang K, Chen J
Information Sources, Risk Perception, and Efficacy Appraisal's Prediction of Engagement in Protective Behaviors Against COVID-19 in China: Repeated Cross-sectional Survey
JMIR Hum Factors 2021;8(1):e23232
URL: http://humanfactors.jmir.org/2021/1/e23232/
doi:10.2196/23232
PMID:33338027

Original Paper

Usability of a Fall Risk mHealth App for People With Multiple Sclerosis: Mixed Methods Study

Katherine Hsieh[1,2*], PhD; Jason Fanning[3*], PhD; Mikaela Frechette[1*], MS; Jacob Sosnoff[1,4*], PhD

[1]Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign, Urbana, IL, United States

[2]Department of Internal Medicine, Section on Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston-Salem, NC, United States

[3]Department of Health and Exercise Science, Wake Forest University, Winston-Salem, NC, United States

[4]Department of Physical Therapy and Rehabilitation Sciences, School of Health Professions, University of Kansas Medical Center, Kansas City, KS, United States

[*]all authors contributed equally

Corresponding Author:
Katherine Hsieh, PhD
Department of Kinesiology and Community Health
University of Illinois at Urbana-Champaign
906 South Goodwin Ave
Urbana, IL, 61820
United States
Phone: 1 2172447006
Email: khsieh@wakehealth.edu

Abstract

Background: Multiple sclerosis (MS) is a chronic, neurodegenerative disease that causes a range of motor, sensory, and cognitive symptoms. Due to these symptoms, people with MS are at a high risk for falls, fall-related injuries, and reductions in quality of life. There is no cure for MS, and managing symptoms and disease progression is important to maintain a high quality of life. Mobile health (mHealth) apps are commonly used by people with MS to help manage their health. However, there are limited health apps for people with MS designed to evaluate fall risk. A fall risk app can increase access to fall risk assessments and improve self-management. When designing mHealth apps, a user-centered approach is critical for improving use and adoption.

Objective: The purpose of this study is to undergo a user-centered approach to test and refine the usability of the app through an iterative design process.

Methods: The fall risk app Steady-MS is an extension of Steady, a fall risk app for older adults. Steady-MS consists of 2 components: a 25-item questionnaire about demographics and MS symptoms and 5 standing balance tasks. Data from the questionnaire and balance tasks were inputted into an algorithm to compute a fall risk score. Two iterations of semistructured interviews (n=5 participants per iteration) were performed to evaluate usability. People with MS used Steady-MS on a smartphone, thinking out loud. Interviews were recorded, transcribed, and developed into codes and themes. People with MS also completed the System Usability Scale.

Results: A total of 3 themes were identified: intuitive navigation, efficiency of use, and perceived value. Overall, the participants found Steady-MS efficient to use and useful to learn their fall risk score. There were challenges related to cognitive overload during the balance tasks. Modifications were made, and after the second iteration, people with MS reported that the app was intuitive and efficient. Average System Usability Scale scores were 95.5 in both iterations, representing *excellent* usability.

Conclusions: Steady-MS is the first mHealth app for people with MS to assess their overall risk of falling and is usable by a subset of people with MS. People with MS found Steady-MS to be usable and useful for understanding their risk of falling. When developing future mHealth apps for people with MS, it is important to prevent cognitive overload through simple and clear instructions and present scores that are understood and interpreted correctly through visuals and text. These findings underscore the importance of user-centered design and provide a foundation for the future development of tools to assess and prevent scalable falls for people with MS. Future steps include understanding the validity of the fall risk algorithm and evaluating the clinical utility of the app.

(JMIR Hum Factors 2021;8(1):e25604) doi:10.2196/25604

KEYWORDS

smartphone; user center design; falls; mobile phone

Introduction

Multiple sclerosis (MS) is a chronic, neurodegenerative disease of the central nervous system (CNS) that affects over a million people in the United States [1]. MS may affect the brain, spinal cord, brainstem, and/or optic nerves and can result in a range of sensory (ie, pain and loss of proprioception), motor (ie, spasticity, muscle weakness, and balance or gait impairments), and/or cognitive (ie, slowed processing speed and memory loss) symptoms [2,3]. Symptoms vary on an individual basis, depending on which areas of the CNS are affected [2]. Furthermore, new symptoms may arise, or current symptoms may worsen throughout the course of the disease [4]. There is currently no cure for MS; however, treatments developed over the last two decades have slowed the disease progression and improved symptoms. Disease-modifying treatments, including injectable and oral drugs, have shown to be beneficial in ameliorating damage to the CNS, and trials using monoclonal antibodies and myelin restoration strategies suggest the potential for novel forms of MS therapy [5]. Although treatments have helped minimize the disease progression, the heterogeneity of MS makes this a complex disease to manage.

Mobile health (mHealth) apps have rapidly evolved in recent years to help individuals track, manage, and treat their health [6]. Due to the complexity of MS, there is increasing use of mHealth apps to support disease monitoring and symptom management [7,8]. More than 85% of people with MS own a mobile device, and 45% of people with MS use an mHealth app to help manage or treat MS [7]. The most common MS apps help with disease management or provide information about MS and MS treatment [9]. Other apps allow people with MS to connect with one another to share information and socialize, and others allow users to track their symptoms, mood, and energy over time [7].

Despite the number of MS-related apps, there are limited health apps developed to evaluate fall risk. Falls are a significant health concern for people with MS, with half of those falling in a 6-month period and up to 50% of falls resulting in an injury [4]. Current fall-related apps for aging and chronic disease populations focus on fall detection [10,11], whereas others measure movement tasks (ie, walking and sit to stand) as a proxy of fall risk [12-14]. Current fall-related apps, however, are not designed for people with MS who have unique risk factors and movement patterns compared with other chronic disease populations. In addition, they did not examine the multiple factors that cause falls in people with MS [4].

Risk factors for falls stem from multiple MS symptoms, including impaired walking and balance, cognitive decline, and fatigue [15]. Although fall risk assessments can be performed clinically, clinicians have time constraints, may not have the necessary equipment, and commonly only assess a single aspect of fall risk, usually asking for previous fall history [16]. Assessing fall risk, however, should include measuring multiple risk factors. Clinical fall risk assessments can include walking and balance tasks such as the Timed up and Go or Short Physical Performance Battery or falls self-efficacy and self-confidence questionnaires [17,18]. A fall risk app incorporating these tasks and measuring multiple risk factors can increase access to fall screening for people with MS and encourage the adoption of fall prevention strategies before a fall occurs. In addition, because MS symptoms fluctuate throughout the course of the disease [4], changes in symptoms lead to changes in fall risk. A fall risk app can help people with MS to measure and track these changes in their homes.

A fall risk mHealth app for people with MS offers access to fall assessment in the home setting, potentially improving fall risk self-management and reducing fall-related injuries. An mHealth app can measure fall risk by leveraging smartphone accelerometry to objectively measure postural control [19] and assess MS symptoms related to falls through self-reported questionnaires. A critical step in the development of an mHealth app is understanding the usability of the app for its intended users [20]. Usability testing ensures that those with MS can easily use and understand an app to improve their overall health. Moreover, a review of MS health apps indicated that most apps do not meet the needs of those with MS because they are not designed for the intended users, leading to poor adherence and use [9]. As people with MS have unique symptoms that may influence their technology use, applying a user-centered approach in the development of health apps can help improve their adoption and use [21]. Therefore, the purpose of this study is to develop a fall risk app for people with MS and to test the usability of the app through an iterative design process. A user-centered approach will improve the development of an app to facilitate the needs of those with MS to increase fall screening and ultimately reduce fall-related injuries [22].

Methods

App Development

This app, Steady-MS, was developed in Android Studio 3.1.2 and was developed as an extension of a validated fall risk app for older adults, Steady [23]. Modifications were made to the questionnaire, balance tasks, and algorithm of Steady to apply specifically to the MS population. Steady-MS consists of 2 components: the first includes 25 questions targeting demographic information and MS symptoms (Multimedia Appendix 1). These questions include age, sex, past history of falls, type of MS, history of MS, the 12-item Multiple Sclerosis Walking Scale (MSWS-12) [24], and the short form of the Activities Balance Confidence Scale (ABC-6; Figure 1) [25]. These questions were specifically chosen because they are associated with falls in people with MS [17,26-28]. The second component, following the 25 questions, is a series of progressive balance tasks, in which the app guides users through 5 progressively difficult standing balance tasks. In the following order, the tasks are as follows: (1) eyes open, (2) eyes closed, (3) semitandem, (4) tandem, and (5) single leg. A text description and image guide users through each task (Figure 2).

Each task takes 30 seconds, beginning with a 5-second countdown and the word *start* and ending with the word *stop*. The phone also vibrates at the start and end of each task. Users were instructed to hold the phone against their chest for the duration of the task to measure their postural sway. These tasks were chosen because worse performance on these tasks is associated with falls in people with MS [29,30]. After each task, users were asked to report if they (1) completed the task, (2) attempted but did not complete, or (3) did not attempt. Steady-MS measures postural sway by measuring acceleration in the mediolateral, anteroposterior, and vertical directions [19]. The Romberg ratio, the ratio between eyes open and closed, of the root mean square acceleration measured and recorded as the increased Romberg ratio, is associated with increased fall risk in people with MS [31]. The number of balance tasks completed, the root mean square Romberg ratio, and the responses to the 25-item questionnaire were inputted into a weighted algorithm and converted into a score ranging from 0 to 100, in which higher scores represent a higher risk for falls.

Figure 1. Screenshots of Steady-MS app asking users to answer 25 questions related to their health, past falls, multiple sclerosis symptoms, and perceived balance. MS: multiple sclerosis.

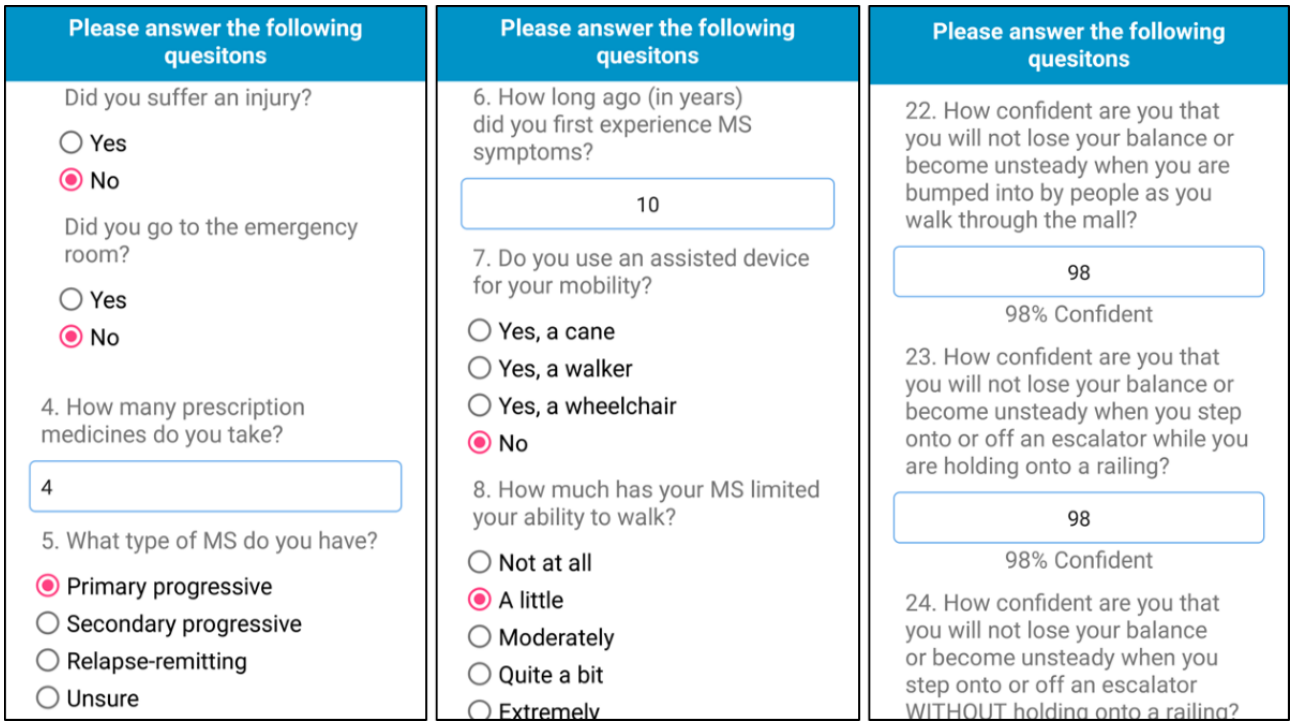
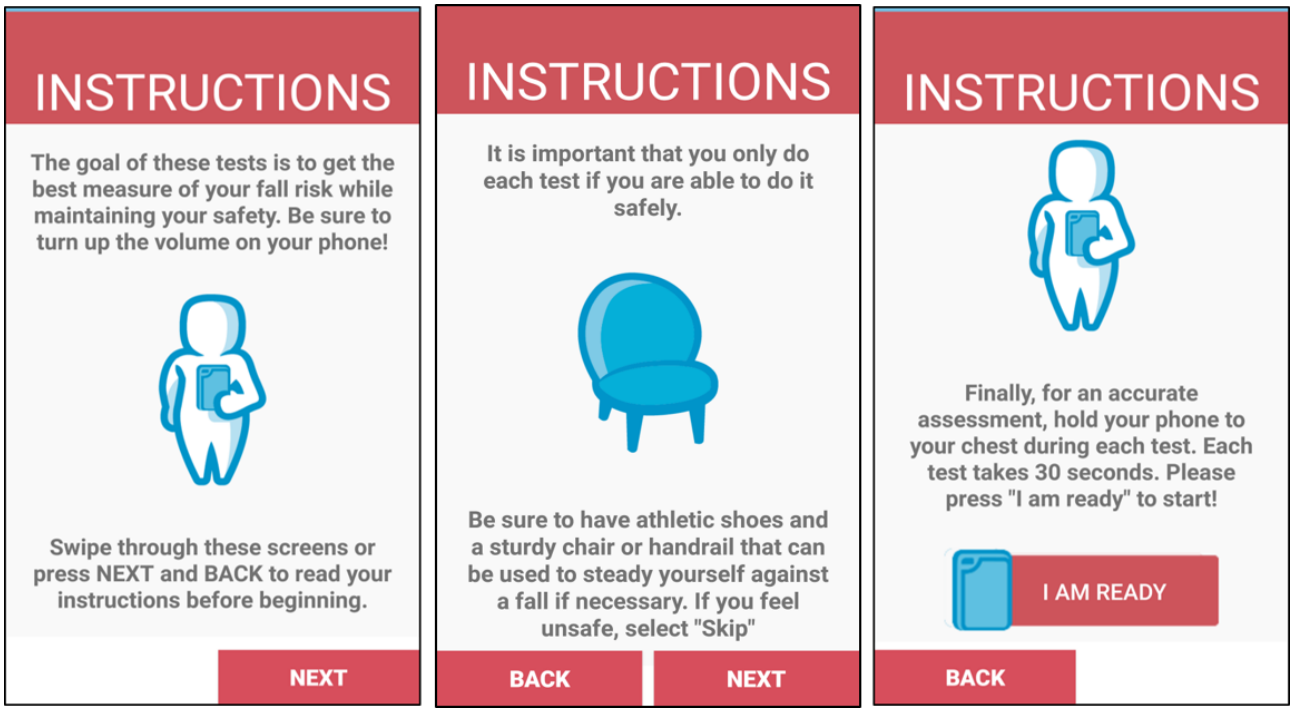


Figure 2. Screenshots of Steady-MS app guiding users to safely perform 5 standing balance tasks while holding the phone against their chest.



Steady-MS was also developed considering common MS symptoms that may influence usability. For instance, fatigue is a common symptom that affects approximately 70% of people with MS [32]. To prevent fatigue, we limited the total number of questions to 25 questions that were needed for the fall risk algorithm and asked only important additional questions (ie, MS duration and type of assisted device) that relate to falls. We also limited the balance tasks to 1 trial of the 5 tasks. Vision impairments are also an MS symptom affecting approximately 30% of people with MS and may influence reading questionnaires and instructions [33]. Therefore, the font size was at least 14, and we emphasized the high contrast between text and background. Cognitive impairment, including reduced processing speed and memory decline, affects between 40% and 70% of people with MS [34]. We aimed to prevent cognitive overload by presenting one set of instructions per screen and maintaining consistency throughout the app.

Participants

A total of 10 people with MS participated in 2 usability rounds. It has been recommended that small groups (n=5) are suitable for identifying usability issues [35]. People with MS (n=5) interacted with Steady-MS and identified usability issues. Using their feedback, we improved the design of the app, and then, another group of people with MS (n=5) interacted with the app to identify any additional usability issues. This iterative design approach centered around the user is most effective for identifying user challenges when developing health apps [21,22]. Inclusion criteria for participants included (1) physician confirmed diagnosis of MS, (2) age 18 years or older, (3) self-reported ability to use a touchscreen device, and (4) ability to stand independently for at least 1 minute. Individuals with neurological disorders other than MS were excluded from the study. All procedures were approved by the Institutional Review Board, and all participants provided informed consent before participation.

Procedures

An iterative design evaluation process of videotaped semistructured interviews was used to determine the optimal usability of Steady-MS. Participants were presented with a smartphone (Samsung Galaxy S6) and asked to open the app and follow all instructions as they completed both the in-app questionnaires and balance tasks. Participants first completed these steps independently, with as little assistance as possible. They then completed the in-app tasks a second time, but this time thinking aloud and narrating their thoughts. They were also encouraged to discuss their likes, dislikes, and recommendations for improvement. After receiving their fall risk score, participants were also asked to identify and draw different graphics of how they wanted to receive their score, such as on a circular chart or linear scale.

Following the semistructured interview, participants completed the Systematic Usability Scale (SUS) to understand the overall usability of the app. The SUS is widely used to quantify the usability of user-machine interfaces, consisting of 10 standard questions on a 5-point Likert scale [36]. The SUS ranges from 0 to 100, with higher scores representing greater usability. Previous work has indicated that the average technology SUS score is 60, and scores of 80 or above indicate that users are more likely to recommend the device to others [37]. Participants also completed the Mobile Device Proficiency Questionnaire (MDPQ) to understand their general proficiency in using mobile devices. The MDPQ ranges from 5 to 40, with higher scores representing greater technological proficiency [38]. Participants then completed the Expanded Disability Status Scale, a self-reported measure of disability that ranges from 0 to 10, with higher scores indicating greater disability [39].

After the first iterative cycle, changes were made to the app design based on the issues identified from the interviews. The second cycle of semistructured interviews was performed on 5 new participants with MS. Owing to COVID-19 restrictions on in-person research, interviews in the second round were performed remotely. The procedures followed the same format as the first round; however, participants were delivered a smartphone with Steady-MS installed, and interviews were conducted over a video call. This format allowed us to understand how Steady-MS is used in the home environment.

Data and Statistical Analysis

All videotapes and field notes taken during the interviews were transcribed verbatim on a computer. Qualitative data from videotapes and field notes were analyzed to develop a coding system using MAXQDA (Version 12.3.3). On the basis of their content, data were assigned codes, and codes with similar content were grouped into themes. The codes and themes were reviewed and discussed by 2 researchers.

Results

Iteration 1

Overview

Participant characteristics are displayed in Table 1. From the semistructured interviews and coding analysis, 3 main themes were identified: (1) intuitive navigation, (2) efficiency of use, and (3) perceived value. Table 2 summarizes the main issues identified from the interviews and the subsequent changes made to Steady-MS.

Table 1. Demographic information of all participants in the first and second iterations.

Variables	Iteration 1	Iteration 2
Age (years), mean (SD)	53.2 (13.1)	54.6 (8.7)
Gender, n (%)		
Female	4 (80)	3 (60)
Male	1 (20)	2 (40)
EDSS[a], median (IQR)	3 (2.5-6)	2.5 (2.5-6)
MS[b] duration (years), mean (SD)	14 (5.9)	16.2 (9.2)
MS type, n (%)		
Primary progressive	1 (20)	0 (0)
Secondary progressive	0 (0)	1 (20)
Relapse remitting	4 (80)	4 (80)
Education, n (%)		
High-school diploma	0 (0)	1 (20)
Associate's degree	2 (40)	1 (20)
Bachelor's degree	3 (60)	1 (20)
Master's degree	0 (0)	2 (40)
Mobile device use, n (%)		
Owns smartphone	5 (100)	4 (80)
Owns tablet	2 (40)	3 (60)
Mobile device proficiency scale, mean (SD)	36.8 (3.3)	38.3 (1.1)

[a]EDSS: Expanded Disability Status Scale.

[b]MS: multiple sclerosis.

Table 2. Summary of the main issues identified in the first round of interviews, sample quotes from each issue, and solutions implemented to improve the app.

Domain and issue	Sample quotes	Solution
Intuitive navigation		
Unclear if eyes are open or closed for balance tests	"I have to keep my eyes closed, don't I?"	Added eyes to icons to depict if eyes are open or closed.
Confusion between semitandem and tandem stances	"Maybe a picture or description because the one that said balance beam made more sense"	Modified pictures to clarify semitandem and tandem stances. Reworded description of each stance.
Reentering ID before balance tests	"I just hit the *Get Started* again?"	After completing *About Me*, users are no longer prompted to reenter their ID.
Redundant option of completing test	"I don't understand *I did not attempt to complete the test* because if you didn't attempt to complete it, why wouldn't you just skip it?"	The *I did not attempt to complete the test* option was removed, as users are able to skip any balance task.
Assisted device use	"This was to think about this as if I'm using my crutch, right?"	Added instructions to answer the activities balance confidence scale as if you were to have your assisted device.
Efficiency of use		
Easy to use	"I find [Steady-MS[a]] easy to use on my own"	No solutions were needed.
Perceived value		
Tracking score over time	If they can learn and improve their score, it would help them feel confident.	No solutions were needed.

[a]MS: multiple sclerosis.

Intuitive Navigation

The most common usability issue during the first iteration was related to intuitive navigation. When participants completed their self-reported questionnaires and moved onto the balance tasks, they were prompted to reenter their ID, a feature that was added to assist in testing several individuals simultaneously. Of the 5 participants, 2 had asked for clarification if they had to reenter their ID or if the app was finished. It was not intuitive for these participants to reenter their ID before moving onto the balance tasks. To address this issue, participants were no longer required to enter their ID to complete the balance tasks. In addition, participants who used an assistive device asked for clarification whether the questionnaires referred to using their assistive device or not. Therefore, for questions such as those from the short form of the ABC, we included instructions regarding assisted device use.

There was also difficulty in navigating through the 5 balance tests. Two of the participants asked for clarification if their eyes were open or closed, whereas 2 different participants performed the semitandem and tandem conditions incorrectly based on observation and video recording from the research staff (Table 3). Although there was a text description instructing each balance stance, these participants reported that they preferred to have a clearer image rather than reading text. In addition, following each balance task, participants were asked to rate if they completed each test with 1 of the 3 options. Of these options, participants reported that the last option, "I did not attempt to complete the test," was not intuitive. Participants indicated that if they were to select this option, they would have chosen to skip the test. To address these issues, we modified the images to indicate that the eyes should be open or closed for each task (Figure 3). We also eliminated the option "I did not attempt to complete the test."

For the final fall risk score, participants also reported that they liked receiving an overall score; however, using a scale to present their score would be the most intuitive to improve their understanding. A total of 3 participants preferred using a horizontal or vertical scale, as opposed to a circular chart. They reported that they understood their score better on a linear scale with *low risk* on one end and *high risk* at the opposite end. Therefore, we added a horizontal scale depicting the user's final fall risk score (Figure 4). The score ranges from 0 to 100, with a green color corresponding to lower fall risk and a red color corresponding to higher fall risk:

> *I'm a visual person, and when I have to read something, I will default to looking at the picture. I mean, I can read an instruction manual all day and, but if you show me a picture or video on how to do it, I'll probably pick it up faster.* [Participant, male, 36 years old]
>
> *You know, like they do on emojis. You just have those little circles for your eyes if they are closed or open. Maybe it's just me, but it's reading all these words or looking at the picture. I could see what I was supposed to do without reading all that.* [Participant, female, 57 years old]
>
> *I don't understand 'I did not attempt to complete the test' because if you didn't attempt to complete it, why wouldn't you just skip it?* [Participant, female, 46 years old]
>
> *I enter my ID again and hit the 'Get Started'?* [Participant, female, 76 years old]

Table 3. Description, order, and number of participants who correctly performed each of the 5 balance tasks in Steady—multiple sclerosis.

Task order	1	2	3	4	5
Visual task	Eyes open	Eyes closed	Eyes open	Eyes open	Eyes open
Feet position	Shoulder width apart	Shoulder width apart	Semi tandem	Tandem	Single leg
Iteration 1 correct performance, n (%)	5 (100)	5 (100)	3 (60)	3 (60)	5 (100)
Iteration 2 correct performance, n (%)	5 (100)	5 (100)	4 (80)	5 (100)	5 (100)

Figure 3. Screenshots of Steady-MS app guiding users through progressive balance tasks. The top panel of screenshots depict the first iteration of images and text, and the bottom panel depicts the second iteration of images and text. Images of eyes and rewording of text were edited to improve clarity and reduce cognitive overload.

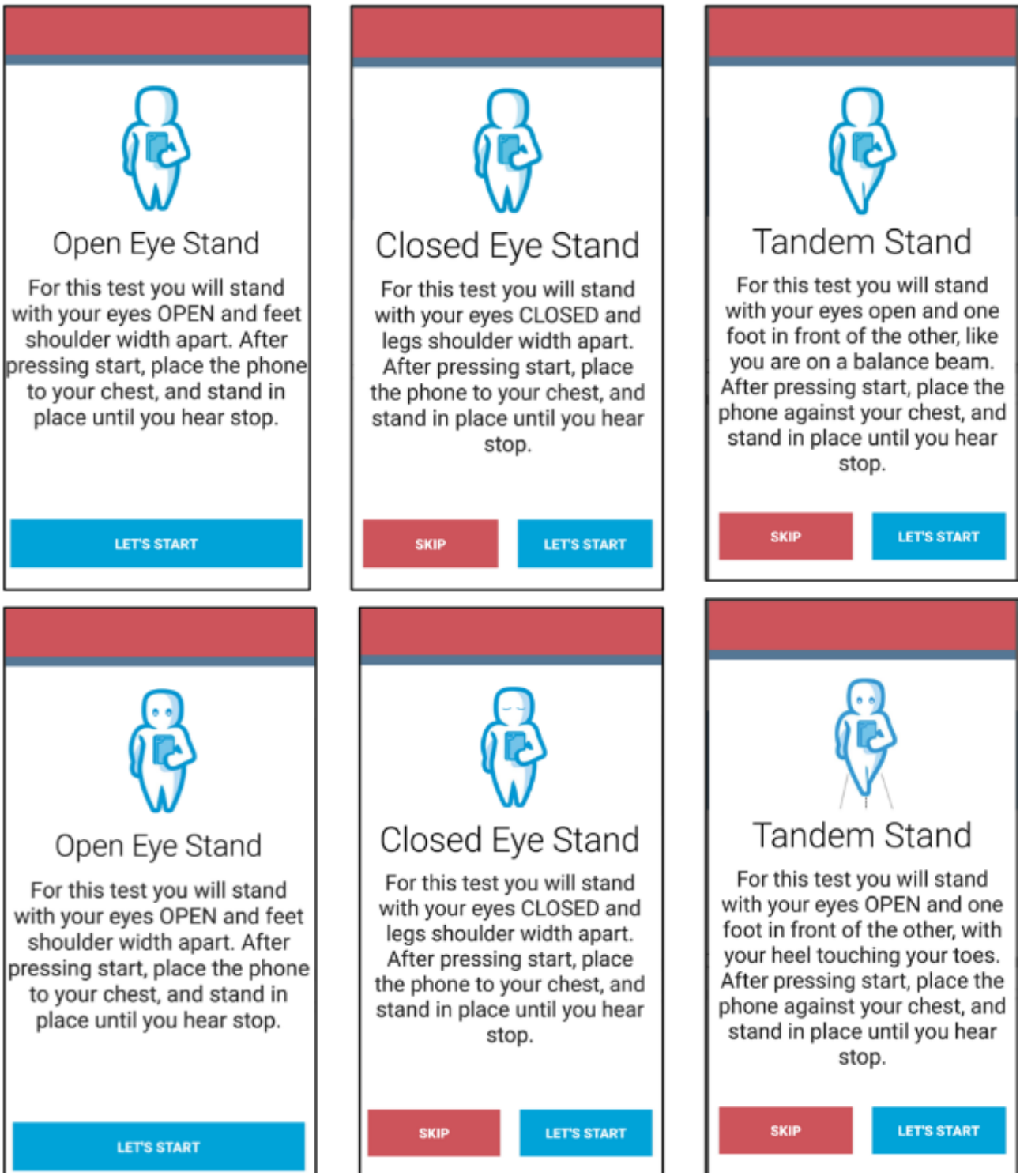
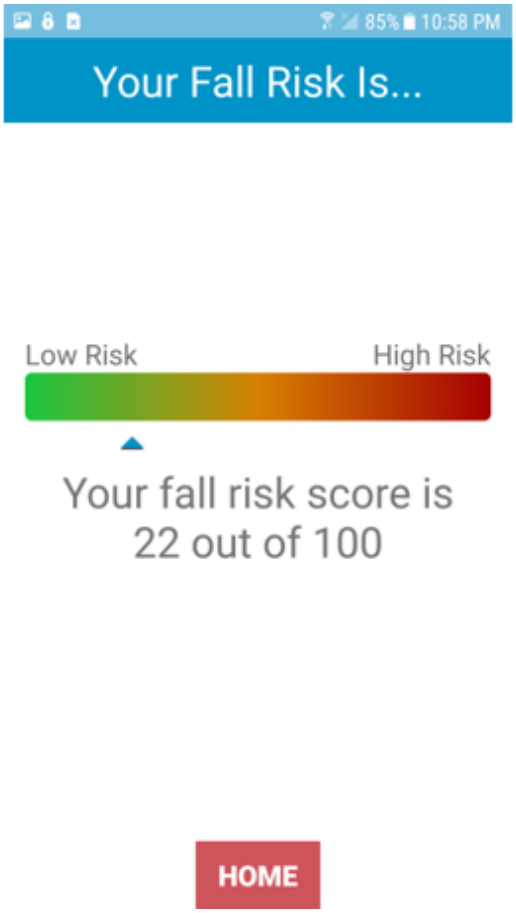


Figure 4. After completing the balance tests, Steady-MS app outputs an overall fall risk score ranging from 0 to 100, with higher scores representing a higher risk of fall.



Efficiency of Use

Overall, all participants found that Steady-MS was efficient and easy to use. Participants reported that the app walks them through each question and balance test and that they could use it independently. One participant reported that the MSWS-12 questionnaire felt redundant; however, none of the participants felt that the total number of questions or number of balance tasks needed to be reduced:

> *I mean, it is pretty easy and seems to walk you through it, in my opinion. It's pretty straight forward.* [Participant, male, 36 years old]
>
> *Everything was quite clear when I was going through that.* [Participant, female, 51 years old]
>
> *I could do that on my own.* [Participant, female, 57 years old]

Perceived Value

The last theme was related to the value of having a fall risk app. All participants reported that having an app would be beneficial for them to understand their risk of falling. Two of the participants found that having a fall risk score can provide confirmation or reassurance in their perceived changes in symptoms, especially during a relapse. These participants said that they would want to use Steady-MS to gauge their changes before seeing a physician. Participants with a higher fall risk found value in learning about their scores; however, they also wanted exercises or other prevention strategies. One participant also reported that she sees value in monitoring her fall risk at home rather than having to travel to a clinician.

Other participants reported that going through the app helped them realize the factors related to falls. One participant, for instance, learned the importance of vision for fall risk and could be more aware of this in the future. Another participant reported that the tandem stance was a balance task that she wanted to improve on:

> *I guess I didn't realize the factors if your eyes open or closed or your stance can increase your fall risk. I guess I can be more conscious about those types of things because it seems to me now with that feedback about my vision, it plays a pretty important role in my balance.* [Participant, female, 57 years old]
>
> *But when I get feeling bad, boy, that number [the fall risk score] shoots up. You know? It's not just my mind, you know, the app kind of confirms it. So maybe I'll use a cane instead.* [Participant, male, 36 years old]
>
> *I like to gauge without having to go all the way to the doctor.* [Participant, female, 46 years old]

Iteration 2

After the second round of interviews, transcript analysis and coding revealed 3 themes related to intuitive navigation, efficiency of use, and perceived value.

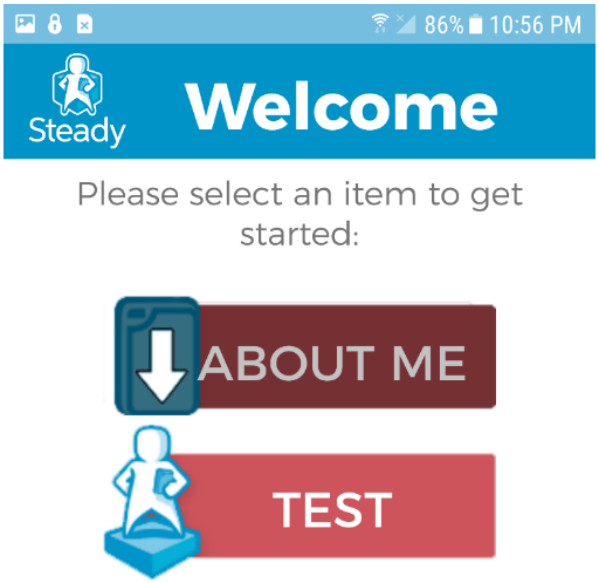
Intuitive Navigation

After modifying Steady-MS, participants in the second round of interviews reported little difficulty navigating through the app. After editing the images and text for the balance tasks, 4 of the 5 participants performed all of the balance tasks correctly based on observation and video analysis. One participant asked for clarification of the semitandem stance to confirm if she was standing correctly. After completing the *About Me* questions, 1 participant returned to the questionnaires again, realized that there were no additional questions, and proceeded to complete the balance questions. To indicate that this section is completed, we dimmed the *About Me* section after users finished the questions (Figure 5). Overall, the participants reported that Steady-MS was intuitive and easy to navigate:

> *I didn't know if there was more about me, like if there were more categories within it. So I chose it again and then I kind of knew enough to be able to scroll through and go back.* [Participant, female, 56 years old]

> *[Referring to the fall risk scale] The green and the red colors [were] pretty self-explanatory to me.* [Participant, female, 53 years old]

Figure 5. After completing the "About Me," this section is dimmed and users are prompted to click on the "Test" section.



Efficiency of Use

Similar to the first round of interviews, participants reported that Steady-MS was efficient and simple to use. They found that navigating through the questions and balance tasks was straightforward. Participants reported that if there was any confusion on the balance tasks or questionnaires, they understood the instructions after rereading a second time. The participants also reported that they could use Steady-MS independently without additional guidance:

> *It seems simple enough to use and I'm not tech savvy as some are. There wasn't anything if I read through it twice I wouldn't understand.* [Participant, male, 61 years old]

> *It's very easy to read. I liked that part, and the contrast is good too. I'm actually reading without my reading glasses, so that's a good sign.* [Participant, female, 53 years old]

> *I thought it was pretty good and straightforward.* [Participant, male, 64 years old]

Perceived Value

All participants reported that Steady-MS can provide many benefits. Participants indicated that the most beneficial component was seeing their fall risk score. For instance, one participant said that when she sees her neurologist, she may be asked to perform static balance tasks but does not receive feedback on her performance. With Steady-MS, she can see a score that gives her measurable feedback. Another participant reported that Steady-MS may be useful in understanding her changes in fall risk with lifestyle changes. Due to COVID-19, her yoga classes have been canceled. She can feel changes in her balance and walking as a result; however, seeing a score to confirm these changes may motivate her to try online yoga:

> *I think it's neat to gauge your risk. Like when I go to the neurologist, she'll have me do stuff, and she'll say hmm or uh huh. And I don't know what any of that means. So it's kind of nice to have it be like, oh, your score is this.* [Participant, female, 39 years old]

> *They're doing a lot of yoga online and whatnot. But we all know we don't do those, or I don't anyway, as much as I would if I were going to class. So it might*

> *be a way for me to say, hey, you need to do a little bit more with your yoga because your balances are getting a little bit more, you know, unstable, I suppose.* [Participant, female, 43 years old]
>
> *You live with yourself all day, every day, and sometimes it's hard to tell if you feel like, you know, like I'm not getting around as well. And if you could look at [the app] and would it show you, oh yeah, it does say I have more of a fall risk.* [Participant, female, 56 years old]

System Usability Scale

In the first iteration, the average SUS score was 95.5, with a standard deviation of 3.3. In the second iteration, the average SUS score was 95.5, with a standard deviation of 2.9. Although the SUS score did not change between iterations, this is likely because of a ceiling effect with a maximum score of 100. These high scores suggest that participants are likely to recommend Steady-MS to others [37].

Discussion

Principal Findings

The purpose of this study is to test and refine the usability of a fall risk health app for individuals with MS through a user-centered design approach. After the first round of usability testing, participants identified issues navigating through the app but reported that it was easy to use and found value in undergoing a fall risk assessment. We modified the app to improve navigation, and after the second round of testing, participants reported that the app was easy to navigate and could use the app on their own. These results, complemented with high SUS scores, suggest that Steady-MS is a usable health app that people with MS can use to self-assess their risk of falling.

Importantly, our results underscore the need for a user-centered design during the development of health apps. Indeed, the main issues identified from semistructured interviews were related to intuitive navigation, and a health app with poor navigation is unlikely to be used. These issues were related to understanding the entire instructions of a balance task (ie, the position of the feet and if eyes are open or closed). Cognitive impairment is a common symptom in people with MS [40], and the instructions for each balance task may result in cognitive overload in people with MS. To reduce cognitive overload, we improved the visuals and text to depict each balance task. Indeed, during the second round of testing, 4 participants completed all balance tasks correctly without asking for clarification. For future developments, it is important to consider the cognitive demands of people with MS to prevent cognitive overload.

Participants in both rounds of testing reported that they found the app clear, simple to use, and useful in learning their risk of falling. This suggests that people with MS can independently use carefully designed health apps such as Steady-MS and learn about their fall risk. Participants also reported that they value receiving a fall risk score because they can identify whether their score improves with exercise or declines with the onset of symptoms. Steady-MS offers the potential for people with MS to self-assess and self-monitor their fall risk using a smartphone. As MS symptoms fluctuate throughout the course of the disease, their risk of falls also changes [4]. Therefore, tracking and monitoring fall risk can help people with MS increase their awareness of their fall risk and take part in prevention strategies before a fall occurs. Unlike traditional fall risk assessments performed in clinics or laboratory-based settings, Steady-MS provides a tool to increase access to fall risk assessments that can be performed at home.

Lessons Learned

When developing future mHealth apps for people with MS, there are important aspects to ensure high usability. First, it is important to prevent cognitive overload in people with MS, as cognitive impairment is a common symptom of MS [40]. Within Steady-MS, cognitive impairment was found when participants were asked to follow 2 separate instructions for a balance task. Using clear visuals and simple text is important to avoid cognitive overload. Second, when presenting a score or number to people with MS, it is important to ensure that the score is easily understandable. Participants in the study reported that they preferred receiving a number because it was measurable, and they could track improvements over time. However, it is important that people with MS accurately interpret scores. When using a scale from 0 to 100, it was important to depict, both visually and through text, that 0 represents low risk and 100 represents high risk. These 2 guidelines can improve the development of future health apps to maximize the usability of people with MS. Third, using a user-centered, iterative approach in designing Steady-MS resulted in users effectively and efficiently understanding new instructions. This approach may also apply to other clinical populations with physical and cognitive impairments when designing a health app.

Limitations

This is the first study to develop and test an MS fall risk mHealth app; however, there are also limitations to this study. The participants in this study had high mobile technology use and scored high on the MDPQ. Those with less technology experience may have additional usability issues that were not identified in this study. However, MS commonly affects younger adults, and more than 80% of people with MS own a smartphone [7]. Therefore, it is likely that a person with MS will already have mobile device experience. Additionally, while Steady-MS measures overall fall risk, it currently does not offer fall prevention strategies. Seeking treatment after understanding individual risk is the next step to prevent falls, and future steps should aim to include tailored fall prevention strategies and understand if people with MS adopt these strategies. Future steps should also understand the validity of the algorithm and display the results of individual components that contribute to fall risk. This may help guide people with MS with specific fall prevention strategies. Saving fall risk scores may also help people with MS monitor their changes over time. In addition, future work should aim to understand how fall risk apps such as Steady-MS can be incorporated into clinical care. Although Steady-MS was designed for use at home, integrating fall risk apps with clinical guidance in a safe manner can increase access to fall prevention strategies. This study of 10 participants is also a small sample size, and future steps should include a larger,

diverse sample to understand usability needs across the heterogeneous MS population. Finally, although participants reported high perceived value in learning about their fall risk score and offered suggestions to improve how the displaying the score, future interviews, and studies should understand how to present individualized fall risk information to prevent negative, unintended consequences.

Conclusions

In conclusion, the purpose of this study is to determine the usability of a fall risk health app for people with MS. After one round of semistructured interviews, we made modifications to improve users' intuitive navigation when answering their health-related questionnaires and performing 5 balance tasks. After a second round of interviews, users reported that the app was straightforward to use and easy to navigate and that they found value in learning about their fall risk. SUS scores averaged 95.5 after the second round of testing, suggesting high usability. These results supported the use of a fall risk app to provide people with MS a tool to self-assess and self-manage their fall risk. Moreover, these results underscored the importance of using a user-centered design approach to identify usability challenges when developing mobile apps for individuals with chronic diseases.

Conflicts of Interest

JJS has ownership in Sosnoff Technologies, LLC. Other authors declared no conflicts of interest.

Multimedia Appendix 1
Steady-MS "About Me" questions.
[DOCX File , 15 KB - humanfactors_v8i1e25604_app1.docx]

References

1. Wallin MT, Culpepper WJ, Campbell JD, Nelson LM, Langer-Gould A, Marrie RA, et al. The prevalence of MS in the United States. Neurology 2019 Feb 15;92(10):e1029-e1040. [doi: 10.1212/wnl.0000000000007035]
2. Filippi M, Bar-Or A, Piehl F, Preziosa P, Solari A, Vukusic S, et al. Multiple sclerosis. Nat Rev Dis Primers 2018 Nov 08;4(1). [doi: 10.1038/s41572-018-0041-4]
3. Frohman EM, Racke MK, Raine CS. Multiple Sclerosis — The Plaque and Its Pathogenesis. N Engl J Med 2006 Mar 02;354(9):942-955. [doi: 10.1056/nejmra052130]
4. Cameron M, Nilsagard Y. Chapter 15 - Balance, Gait,Falls in Multiple Sclerosis. In Day BL. Lord SR (eds): Handbook of Clinical Neurology. Elsevier - 2018;237:250. [doi: 10.1016/b978-0-444-63916-5.00015-x]
5. Gholamzad M, Ebtekar M, Ardestani MS, Azimi M, Mahmodi Z, Mousavi MJ, et al. A comprehensive review on the treatment approaches of multiple sclerosis: currently and in the future. Inflamm. Res 2018 Sep 3;68(1):25-38. [doi: 10.1007/s00011-018-1185-0]
6. 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]
7. Marrie RA, Leung S, Tyry T, Cutter GR, Fox R, Salter A. Use of eHealth and mHealth technology by persons with multiple sclerosis. Multiple Sclerosis and Related Disorders 2019 Jan;27:13-19. [doi: 10.1016/j.msard.2018.09.036]
8. Lowe-Strong A, McCullagh PJ. Monitoring of symptoms and interventions associated with multiple sclerosis. Stud Health Technol Inform 2005;117:223-228. [Medline: 16282674]
9. Giunti G, Guisado Fernández E, Dorronzoro Zubiete E, Rivera Romero O. Supply and Demand in mHealth Apps for Persons With Multiple Sclerosis: Systematic Search in App Stores and Scoping Literature Review. JMIR Mhealth Uhealth 2018 May 23;6(5):e10512. [doi: 10.2196/10512]
10. Casilari E, Santoyo-Ramón JA, Cano-García JM. Analysis of a Smartphone-Based Architecture with Multiple Mobility Sensors for Fall Detection. PLoS ONE 2016 Dec 8;11(12):e0168069. [doi: 10.1371/journal.pone.0168069]
11. Ning Y, Hu S, Nie X. Real-time Action Recognition and Fall Detection Based on Smartphone. 2018 Jul 21 Presented at: Annu Int Conf IEEE Eng Med Biol Soc; 2018; Honolulu, HI p. 4418-4422. [doi: 10.1109/embc.2018.8513314]
12. Cerrito A, Bichsel L, Radlinger L, Schmid S. Reliability and validity of a smartphone-based application for the quantification of the sit-to-stand movement in healthy seniors. Gait & Posture 2015 Feb;41(2):409-413. [doi: 10.1016/j.gaitpost.2014.11.001]
13. Rasche P, Mertens A, Bröhl C, Theis S, Seinsch T, Wille M, et al. The "Aachen fall prevention App" – a Smartphone application app for the self-assessment of elderly patients at risk for ground level falls. Patient Saf Surg 2017 May 8;11(1). [doi: 10.1186/s13037-017-0130-4]
14. Roeing KL, Hsieh KL, Sosnoff JJ. A systematic review of balance and fall risk assessments with mobile phone technology. Archives of Gerontology and Geriatrics 2017 Nov;73:222-226. [doi: 10.1016/j.archger.2017.08.002]
15. Giannì C, Prosperini L, Jonsdottir J, Cattaneo D. A systematic review of factors associated with accidental falls in people with multiple sclerosis: a meta-analytic approach. Clin Rehabil 2014 Feb 25;28(7):704-716. [doi: 10.1177/0269215513517575]
16. Rae-Grant A, Bennett A, Sanders AE, Phipps M, Cheng E, Bever C. Quality improvement in neurology: Multiple sclerosis quality measures: Table. Neurology 2015 Sep 02;85(21):1904-1908. [doi: 10.1212/wnl.0000000000001965]

17. Finlayson ML, Peterson EW, Cho CC. Risk Factors for Falling Among People Aged 45 to 90 Years With Multiple Sclerosis. Archives of Physical Medicine and Rehabilitation 2006 Sep;87(9):1274-1279. [doi: 10.1016/j.apmr.2006.06.002]
18. Gunn HJ, Newell P, Haas B, Marsden JF, Freeman JA. Identification of risk factors for falls in multiple sclerosis: a systematic review and meta-analysis. Phys Ther 2013 Apr;93(4):504-513. [doi: 10.2522/ptj.20120231] [Medline: 23237970]
19. Hsieh KL, Sosnoff JJ. Smartphone accelerometry to assess postural control in individuals with multiple sclerosis. Gait & Posture 2021 Feb;84:114-119. [doi: 10.1016/j.gaitpost.2020.11.011]
20. Broekhuis M, van Velsen L, Hermens H. Assessing usability of eHealth technology: A comparison of usability benchmarking instruments. International Journal of Medical Informatics 2019 Aug;128:24-31. [doi: 10.1016/j.ijmedinf.2019.05.001]
21. McCurdie T, Taneva S, Casselman M, Yeung M, McDaniel C, Ho W, et al. mHealth Consumer Apps: The Case for User-Centered Design. Biomedical Instrumentation & Technology 2012 Jan 01;46(s2):49-56. [doi: 10.2345/0899-8205-46.s2.49]
22. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). Journal of Biomedical Informatics 2016 Apr;60:243-251. [doi: 10.1016/j.jbi.2016.02.002]
23. Hsieh KL, Fanning JT, Rogers WA, Wood TA, Sosnoff JJ. A Fall Risk mHealth App for Older Adults: Development and Usability Study. JMIR Aging 2018 Nov 20;1(2):e11569. [doi: 10.2196/11569]
24. Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ. Measuring the impact of MS on walking ability: The 12-Item MS Walking Scale (MSWS-12). Neurology 2003 Jan 14;60(1):31-36. [doi: 10.1212/wnl.60.1.31]
25. Wood T, Wajda D, Sosnoff J. Use of a Short Version of the Activities-specific Balance Confidence Scale in Multiple Sclerosis. International journal of MS care 2019 Jan 01:15-21. [doi: 10.7224/1537-2073.2017-047]
26. Cameron M, Thielman E, Mazumder R, Bourdette D. MS International. Predicting Falls in People with Multiple Sclerosis: Fall History Is As Accurate As More Complex Measures. Mult Scler Int 2013; 2013:496325.
27. Cattaneo D, De Nuzzo C, Fascia T, Macalli M, Pisoni I, Cardini R. Risks of falls in subjects with multiple sclerosis. Archives of Physical Medicine and Rehabilitation 2002 Jun;83(6):864-867. [doi: 10.1053/apmr.2002.32825]
28. Nilsagård Y, Gunn H, Freeman J, Hoang P, Lord S, Mazumder R, et al. Falls in people with MS—an individual data meta-analysis from studies from Australia, Sweden, United Kingdom and the United States. Mult Scler 2014 Jun 16;21(1):92-100. [doi: 10.1177/1352458514538884]
29. Comber L, Sosnoff JJ, Galvin R, Coote S. Postural control deficits in people with Multiple Sclerosis: A systematic review and meta-analysis. Gait & Posture 2018 Mar;61:445-452. [doi: 10.1016/j.gaitpost.2018.02.018]
30. Prosperini L, Fortuna D, Giannì C, Leonardi L, Pozzilli C. The Diagnostic Accuracy of Static Posturography in Predicting Accidental Falls in People With Multiple Sclerosis. Neurorehabil Neural Repair 2012 May 15;27(1):45-52. [doi: 10.1177/1545968312445638]
31. Kalron A. The Romberg ratio in people with multiple sclerosis. Gait & Posture 2017 May;54:209-213. [doi: 10.1016/j.gaitpost.2017.03.016]
32. Lerdal A, Celius E, Krupp L, Dahl A. A Prospective Study of Patterns of Fatigue in Multiple Sclerosis. Eur J Neurol 2007;14:1338-1343. [doi: 10.1111/j.1468-1331.2007.01974.x]
33. Balcer LJ, Miller DH, Reingold SC, Cohen JA. Vision and vision-related outcome measures in multiple sclerosis. Brain 2014 Nov 28;138(1):11-27. [doi: 10.1093/brain/awu335]
34. Langdon DW. Cognition in multiple sclerosis. Current Opinion in Neurology 2011;24(3):244-249. [doi: 10.1097/WCO.0b013e328346a43b] [Medline: 21519256]
35. Nielsen J. How many test users in a usability study. Nielsen Norman Group 2012;4.
36. Brooke J. SUS-A Quick and Dirty Usability Scale. In: Usability Evaluation In Industry. Bristol, PA: Taylor & Francis; 1996:189-196.
37. Sauro J. Measuring Usability with the System Usability Scale. 2011. URL: https://measuringu.com/sus/
38. Roque NA, Boot WR. A New Tool for Assessing Mobile Device Proficiency in Older Adults: The Mobile Device Proficiency Questionnaire. J Appl Gerontol 2016 Apr 11;37(2):131-156. [doi: 10.1177/0733464816642582]
39. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology 1983 Nov 01;33(11):1444-1444. [doi: 10.1212/wnl.33.11.1444]
40. Sumowski JF, Benedict R, Enzinger C, Filippi M, Geurts JJ, Hamalainen P, et al. Cognition in multiple sclerosis. Neurology 2018 Jan 17;90(6):278-288. [doi: 10.1212/wnl.0000000000004977]

Abbreviations

ABC: Activities Balance Confidence Scale
CNS: central nervous system
MDPQ: Mobile Device Proficiency Questionnaire
mHealth: mobile health
MS: multiple sclerosis
MSWS-12: The 12-item Multiple Sclerosis Walking Scale

SUS: System Usability Scale

Edited by A Kushniruk; submitted 08.11.20; peer-reviewed by L Omberg, R Bergquist, C Blair; comments to author 23.12.20; revised version received 20.01.21; accepted 22.02.21; published 22.03.21.

Please cite as:
Hsieh K, Fanning J, Frechette M, Sosnoff J
Usability of a Fall Risk mHealth App for People With Multiple Sclerosis: Mixed Methods Study
JMIR Hum Factors 2021;8(1):e25604
URL: https://humanfactors.jmir.org/2021/1/e25604
doi:10.2196/25604
PMID:33749609

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org